

基于隐马尔可夫模型的语音识别技术在载人航天器上的应用

李皖玲, 梁吴迪, 张天湘

(中国空间技术研究院 载人航天总体部, 北京 100094)

摘要: 利用嵌入式平台构建了基于隐马尔可夫模型 (Hidden Markov Model, HMM) 技术的载人航天器语音识别系统, 并针对载人航天器特殊的噪声背景和声学训练, 对传统语音端点检测算法进行了改进。通过试验在嵌入式运行环境下实现了指令发送、文字输入等功能, 证实该系统适应航天员的口音和发音习惯, 可对连贯自然发音方式的命令进行识别, 能够满足航天员及任务使用需求。

关键词: 载人航天器; 语音识别; 隐马尔可夫模型; 硬件设计; 软件设计

中图分类号: TP912.34

文献标志码: A

文章编号: 1673-1379(2013)04-0441-05

DOI: 10.3969/j.issn.1673-1379.2013.04.020

0 引言

语音识别即利用计算机解读人的语音, 并转换为相应的文本信息及指令, 用于信息输入及系统控制, 是当前人机交互研究领域的热点之一^[1]。

目前, 载人航天器控制的输入方式仍然以手动操作为主, 但随着有人参与任务的复杂性增加, 座舱操作面积与手动操作需求之间的矛盾日益突出, 若想完成更为复杂的任务, 仅用手动方式难以实现。近几年, 随着语音识别技术在各个领域的应用推广, 语音输入技术在飞行器控制上的应用研究正逐渐受到重视^[3-4]。

美国空军研究实验室 (Air Force Research Laboratory) 针对无人机的控制, 分别以手动输入与语音输入模式进行了大量的控制指令输入实验, 实验数据显示语音输入具有明显的优势^[2]: 执行效率高, 任务完成时间可缩短近 40%; 错误率低, 执行错误率可降低 1/3; 交互效果好, 更符合人类交流习惯。

语音识别根据所处理语音对象的不同, 可分别归类为: 特定人或非特定人; 小、中、大词汇量; 孤立词语音识别或连续语音识别。本文研究针对载人航天器控制输入 (固定模式航天员操作口令), 其语音识别系统是特定人、特定小词汇量的连续语

音识别系统。

1 隐马尔可夫模型 (HMM)

隐马尔可夫模型 (Hidden Markov Model, HMM) 是一种统计模型, 其语法已被语音识别领域广泛采用, 是当前最有效的语音识别方法^[4]。

HMM 是对信号的时间序列建立统计模型, 并将其作为数学上的双重随机过程, 原理如图 1 所示, 分为 Markov 链 (π, A) 和随机过程 (B) 两部分。Markov 链的输出为状态序列 S ; 随机过程产生的输出为观察值序列 O ^[5-6]。

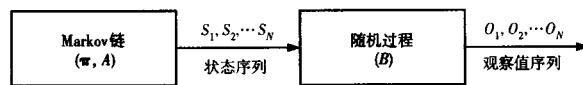


图 1 隐马尔可夫模型原理

Fig. 1 Principle of Hidden Markov Model

记一个 HMM 为

$$\lambda = (\pi, A, B), \quad (1)$$

其中: π 为起始状态概率, 用于描述观察值序列 O 在 $t=1$ 时刻所处状态。记 t 时刻的概率为 q_t , 即

$$\pi = \{\pi_i\}, \pi_i = P(q_1 = S_i), i = 1, 2, \dots, N, \quad (2)$$

且满足

$$\sum_{i=1}^N \pi_i = 1. \quad (3)$$

收稿日期: 2013-02-05; 修回日期: 2013-07-22

基金项目: 中国空间技术研究院载人航天总体部自主研发课题

作者简介: 李皖玲 (1981—), 女, 硕士学位, 主要从事载人飞船仪表与照明分系统设计工作。Email: liwanling66@gmail.com。

A 为状态转移概率。即

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i), \quad (4)$$

且满足

$$\sum_{j=1}^N a_{ij} = 1. \quad (5)$$

B 为观察值序列 O 的任一观察值在空间分布的概率,分为离散型和连续型两类。离散 HMM 情况下,

$$B = \{b_j(k), j = 1, 2, \dots, N; k = 1, 2, \dots, M\}, \quad (6)$$

其中 M 为编码符号集中符号的总数,且满足

$$\sum_{k=1}^M b_j(k) = 1. \quad (7)$$

连续 HMM 情况下,

$$B = \{b_j(O), j = 1, 2, \dots, N\}, \quad (8)$$

其中 O 为观察值矢量空间的任一矢量,且满足

$$\int_{\Omega_j} b_j(O) dO = 1, \quad (9)$$

其中 Ω_j 表示第 j 状态的观察概率空间。

人的言语过程即双重随机过程,语音信号是可观测的时变序列,是根据语法知识和言语需要(不可观测的状态)产生的音素的数据流,因此语言过程可用 HMM 模型进行描述。

2 基于 HMM 技术的语音识别算法

基于 HMM 技术的语音识别算法可以描述为:在给定输入语音观测序列的条件下,在声学模型和语言模型的指导下,寻找一条最佳的解码路径,使得该路径下的词序列产生输入语音观察值序列的概率最大^[8]。

典型的语音识别算法包含有特征提取(频谱分析)、模式匹配和模型库等 3 个基本单元,系统结构如图 2 所示^[9-10]。

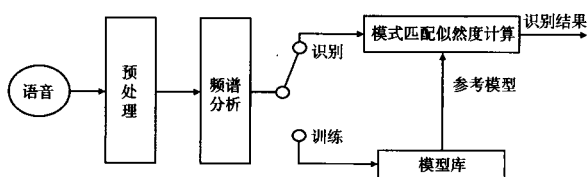


图 2 语音识别算法系统结构

Fig. 2 Block diagram of speech recognition algorithm

预处理单元对输入的语音信号进行处理,执行

A/D 转换,并滤除信号中的背景噪声及不重要的信息;同时进行语音信号的端点检测、预加重、噪声抵消,以弥补语音信号高频部分大约每 10 倍频程衰减 20 dB 的损失特性。

频谱分析单元对预处理后的数字信号进行分析处理,提取信号中可反映语音信号声学特征的特征,并形成矢量。

模式匹配与识别决策则是依据一定准则及先验知识(语法规则、词法规则和语义规则),将频谱分析形成的语音指令与模型库中某一参考模式进行比较,获得最佳匹配并将候选结果输出。

训练阶段是对已输入语音进行预处理与频谱分析,得到语音的特征矢量序列,按照准则建立该模式特征的模型库,或是对模型库中的特征参数进行修正。

3 载人航天器语音识别系统实现方案

基于 HMM 模型和语音识别系统,利用 Visual C++ 开发平台,构建了载人航天器语音识别系统,该系统有语音训练和语音识别 2 种工作模式。

3.1 硬件设计

载人航天器语音识别系统以嵌入式处理器为核心,输入输出设备包括话筒、扬声器和显示器,软件程序存储在 EEPROM 中,FLASH 存储模型库。系统硬件架构如图 3 所示。

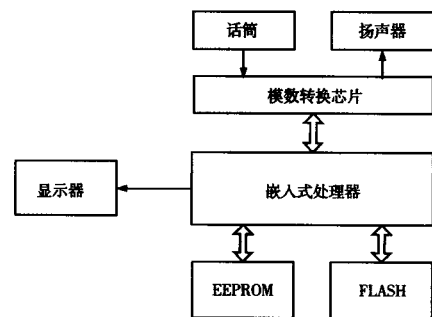


图 3 语音识别系统硬件架构

Fig. 3 Hardware structure of speech recognition system

语音训练模式下,从话筒输入的语音信号经 A/D 转换后送入核心处理器,由处理器利用 HMM 技术进行分析,并将分析结果存储在 FLASH 中。

语音识别模式下,从话筒输入的语音信号经 A/D 转换后送入核心处理器,处理器依照基于 HMM

技术的软件系统，将处理过的信号与存储在 FLASH 模型库中的参考模式进行对比，找出最相近的指令作为识别结果输出给语音识别的应用程序。

3.2 软件设计

载人航天器语音识别系统的软件主要由语音识别程序和交互应用程序构成。语音识别程序实现语音识别及训练功能，包含语音信号处理模块及 HMM 识别模块。交互应用程序用于完成用户指定功能，不需要关心语音识别过程和算法，只需调用语音识别程序的识别结果。软件架构如图 4 所示。

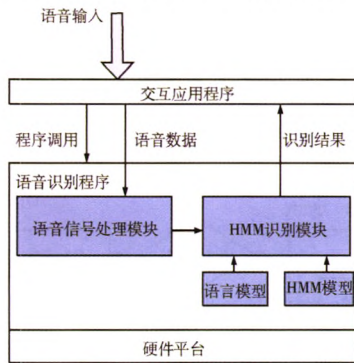


图 4 语音识别系统软件架构

Fig. 4 Software structure of speech recognition system

语音识别程序中的算法流程如图 5 所示，程序流程如图 6 所示。

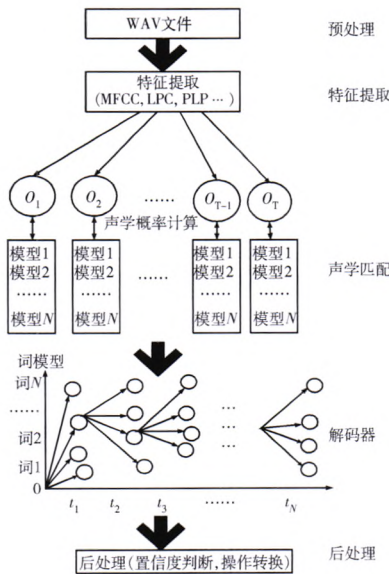


图 5 语音识别算法流程

Fig. 5 Process chart of speech recognition algorithm

在语音识别程序中，系统初始化完成后，通过人机交互界面提示选择进入训练模式或识别模式；软件接收 A/D 转换后的语音信号，调用语音识别算法模块，实现模型参数存储或语音识别输出。

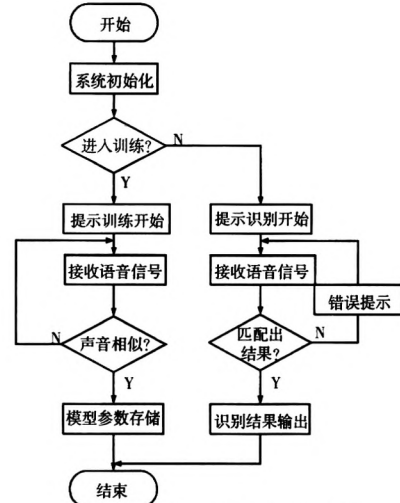


图 6 语音识别程序软件流程

Fig. 6 Flow chart of speech recognition software

4 语音端点检测算法改进

载人航天器语音识别系统实际应用环境中的噪声会对系统的识别效果产生影响，需要分析噪声源，对语音端点检测算法进行改进，以提高系统的识别率，降低拒识率，缩短响应时间。

4.1 针对轨道运行段噪声环境的处理

载人航天器轨道运行段的主要噪声源来自舱内生命保障系统，其总声压级约为 60~75 dB^[11-12]，背景噪声频谱如图 7 所示。

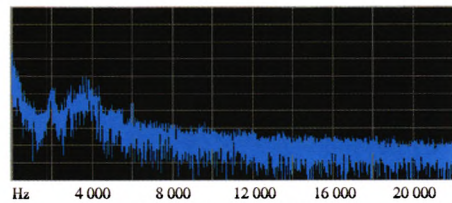


图 7 背景噪声频谱

Fig. 7 Frequency spectrum of background noise

由图 7 可知，背景噪声主要集中在 200~500 Hz、2000~2100 Hz 和 3800~4100 Hz 这 3 个窄频带内。

因噪声对语音信号各频段的影响不同，带噪语音信号各频段的信噪比也不同。在无噪声语音训练前提下，信噪比低的缺失特性对识别起负作用。在语音识别系统中，可采用缺失特性重建方法对噪声进行处理，其基本思路是语音信号各频段之间存在相关性，利用相关性可由高信噪比语音信号的可靠特性估计低信噪比语音信号的缺失特性。处理流程如下：

1) 对每一帧输入的信号利用语音判别器判别其是噪声信号还是语音信号，如果是噪声信号，则

直接过滤掉;

2) 经过处理的语音信号输入至低频滤波器, 过滤掉低频噪声信号;

3) 提取语音信号的频谱特征, 分析剩余的噪声主要集中在频率点;

4) 去除被噪声污染的频率点的频谱特征, 利用其相邻的未被污染的语音帧的频谱重新估计频谱特征, 替代被污染的频谱特征。

4.2 针对大过载噪声环境的处理

载人航天器在发射和返回时产生的过载会使航天员的发声系统变异(如面部变形扭曲导致口腔发音变化), 发出的语音指令与正常情况下的不一样, 不能与正常模型相匹配, 导致语音识别率的下降。

本系统的解决方法是在过载条件下对声音样

本进行采集(例如在 2.5g、3g、3.5g、4g 等情况下进行采集; 如果条件允许, 继续在 4.5g、5g、6g 等更大过载情况下进行采集), 并通过对这些样本信号的分析, 修正非过载条件下的语音模型, 以模型自适应的方法得到过载条件下的语音模型。在过载条件下, 用过载语音模型替换非过载语音模型进行识别, 提高大过载环境下的识别率。

5 载人航天器语音识别系统试验验证

5.1 语音识别试验方法

载人航天器语音识别系统由麦克风、语音识别嵌入式系统和控制按钮组成。受试者操作控制按钮, 语音通过麦克风输入, 嵌入式系统作为语音识别设备。试验流程如图 8 所示。

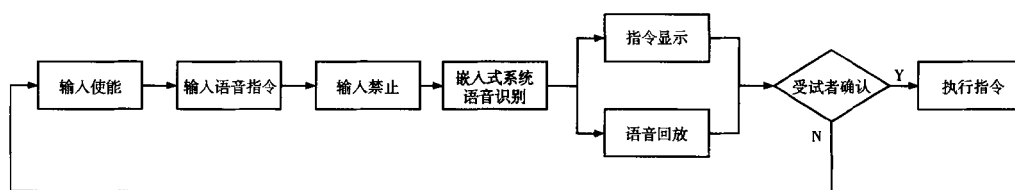


图 8 语音识别试验流程

Fig. 8 Flow chart of speech recognition testing

1) 受试者将系统设置为输入使能后, 通过麦克风进行语音指令的输入, 输入完毕后再将系统设置为输入禁止。

2) 嵌入式系统中的语音识别程序对输入的语音指令信号进行识别, 产生识别结果, 在交互界面显示指令, 同时将识别结果转化成语音信号通过扬声器回放。

3) 受试者操作“是”或“否”键控制按钮表示是否确认本次识别结果, “是”则由交互应用程

序执行与识别结果相应的指令, 完成一次语音识别试验; “否”则返回步骤 1), 重新进行语音输入。

4) 受试者重复进行语音识别试验。

5.2 识别结果

语音识别系统主要实现轨实时语音识别处理, 其可实现的性能依赖于系统配置的硬件资源。表 1 为不同的硬件资源配置下, 基于 HMM 技术的载人航天器语音识别系统所能达到的识别结果。

表 1 语音识别结果
Table 1 Results of speech recognition

核心处理器		内存/M	识别结果				
处理速度/MHz	位数/bit		噪声环境	识别语音数/条	识别率/%	拒识率/%	响应时间/s
200	16	16	不考虑	50 以内	>99	>70	<0.3
300	16	24	考虑				
400	16	30		500 以内	>97		
500	16	45					

由表 1 可见, 在嵌入式运行环境下, 核心处理器速度在 200 MHz 以上, 即可实现对连贯自然发音方式的命令进行识别。随着核心处理器能力的提升, 即使考虑航天器噪声背景下, 识别能力也将大幅提高, 识别率达 97% 以上。

6 结束语

本文基于 HMM 技术构建了载人航天器语音识别系统, 通过对传统语音端点检测算法改进, 最终在嵌入式运行环境实现了噪声环境下的语音识别功能, 通过试验证实该系统可适应航天员的口音

和发音习惯,实现对连贯自然发音方式的命令进行识别,能够满足航天员及任务使用需求。

目前,载人航天器语音识别技术可实现小词汇量、非特定人的孤立词语音识别,但与真正的人机自由交互还有相当的差距。随着科学技术水平的发展,后续载人航天器的语音识别技术需从孤立词的语音识别进展为大词汇量的连续不规则语音识别,将概率模型转为以语言学为基础的文法模型,从而使语音识别技术真正成为未来载人航天器人机交互的基本手段。

参考文献 (References)

- [1] 王炳锡, 屈丹, 彭煊. 实用语音识别基础[M]. 北京: 国防工业出版社, 2005: 2-4
- [2] Williamson D T. Performance assessment of dynaspeak speech recognition system on in-flight databases[G]. US Air Force Research Laboratory, 2004: 37-38
- [3] Draper M, Calhoun G, Ruff H, et al. Manual versus speech input for unmanned aerial vehicle control station operations[C]//Proceedings of the Human Factors and Ergonomics Society, 2003-10: 109-113
- [4] 郭小明. 飞机座舱显示控制工效学研究近况[J]. 人机工效学, 2001, 7(4): 34-37
Guo Xiaoming. Study of display-control efficiency on flight cabin[J]. Chinese Journal of Ergonomics, 2001, 7(4): 34-37
- [5] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2001: 68-69
- [6] Williamson D T, Barry T P, Liggett K K. Flight test results of ITT VRS-1290 in NASA OV-10[C]//Proceedings of AVIOS'96 15th Annual International Voice Technologies Conference, 1996: 33-42
- [7] 梁向东. 语音识别技术应用于航空控制系统[J]. 洪都科技, 2003(3): 2-3
Liang Xiangdong. Application of SR technique to aviation control system[J]. Hongdu Science and Technology, 2003(3): 2-3
- [8] Cerf P L, Comperolle D V. A new variable frame rate analysis method for speech recognition[J]. IEEE Signal Processing Letters, 1994, 12(12): 185-187
- [9] Fu Qiang, Yi Kechu. Bark wavelet transform of speech and it's application in speech recognition[J]. Journal of Electronics, 2000, 10(28): 26
- [10] 朱民雄, 闻新, 黄健群, 等. 计算机语音技术[M]. 北京航空航天大学出版社, 2001: 164-167
- [11] 李锋, 刘小勇, 鲍福廷, 等. 火箭发动机噪声测试及预估研究[J]. 电声技术, 2009, 33(9): 53-55
Li Feng, Liu Xiaoyong, Bao Futing, et al. Research on noise measuring and prediction of rocket engine[J]. Audio Engineering, 2009, 33(9): 53-55
- [12] 噪声困扰航天员[J]. 太空探索, 2002(2): 14
Noise interference on astronauts[J]. Space Exploration. 2002(2): 14

Spacecraft speech recognition based on Hidden Markov Model (HMM) technology

Li Wanling, Liang Wudi, Zhang Tianxiang

(Beijing Institute of Manned Space System Engineering, China Academy of Space Technology, Beijing 100094, China)

Abstract: This paper reviews the theory of the Hidden Markov Model (HMM) and the speech recognition system. By using the embedded platform, the manned -spacecraft speech recognition system is built based on the HMM technology. To meet complex requirements of future spacecraft for speech recognition, the specifications and rules for the speech recognition are optimized, with an improvement of the accuracy rate over the traditional voice activity detection algorithm. The application of this continuous speech recognition system on manned-spacecraft shows good results for the function implementation of command sending and character input.

Key words: manned-spacecraft; speech recognition; Hidden Markov Model; hardware design; software design

(编辑: 冯妍)