

基于中文语音识别技术的舰艇指挥训练系统的研究

景春进¹, 陈东东², 周琳琦¹

(1. 中国卫星海上测控部, 江苏 江阴 214431; 2. 江阴职业技术学院, 江苏 江阴 214405)

摘要: 针对舰艇指挥训练系统的特点, 提出了一种利用语音识别技术提高其训练效率的方法; 首先分析了舰艇指挥指令的语言特点, 然后研究了基于 Sphinx 平台的汉语连续语音识别的相关问题, 包括声学模型的训练、语言模型的训练及语音识别引擎等; 最后设计并实现了一个非特定人, 中等专用词汇量的连续汉语语音识别系统; 实验采用了一定数量的数字和专用词汇进行验证, 结果表明, 经过声学模型训练后, 该系统的识别率有较大提高; 该方法对提高舰艇指挥训练系统的自动化水平具有一定的指导意义。

关键词: 汉语语音识别; 指挥指令; 语音特征提取; 声学模型; 语言模型

Research on Vessel Command Training System Based on Chinese Speech Recognition Method

Jing Chunjin¹, Chen Dongdong², Zhou Linqi¹

(1. China satellite Maritime Tracking and control department, Jiangyin 214431, China;

2. Jiangyin Institute of Vocation Technology, Jiangyin 214405, China)

Abstract: A method which will improve the training efficiency is proposed in this paper with the help of speech recognition technology due to the features of naval vessel command system. In the first place, the linguistic features of vessel command instruction are analyzed. And then the relevant problems about continuous Chinese speech recognition are studied on the basis of Sphinx platform, which includes the training of acoustic model, the one of language model, the engine of speech recognition, and so forth. Based on the analyses a non-specific, moderate special Chinese vocabulary continuous speech recognition system is designed and finally comes into being. Experiments use a certain amount of numbers and specialized vocabulary, The results indicate that the system has a high recognition rate after the training of acoustic model. The method of this paper has important guidance for improving the automatic level of the system.

Keywords: Chinese speech recognition; command and control instruction; speech feature extraction; acoustic model; language model

0 引言

舰艇指挥调度系统是影响舰艇作战效能的关键环节, 舰艇大型化和多功能化的发展使得指挥调度系统日趋复杂。为了培养各级指挥调度人员, 保证指挥调度系统在作战时发挥最佳作用, 必须对系统使用人员进行长期的模拟实战演练, 就需要舰艇指挥训练系统的支持, 通过利用计算机仿真技术模拟舰艇的整个指挥调度系统, 对人员进行训练。传统的舰艇指挥训练系统中需要每个调度席位都必须配备相应的人员, 由教辅舰员在相应的岗位发出各类指挥调度命令, 其他受训舰员进行语音应答和相关操作。随着训练人数和次数的增加, 教辅舰员工作量大增, 特别是大型舰艇, 参训人员多, 这种模拟训练方式会大大降低训练效率。而利用语音合成和识别技术创建模拟指挥调度席位, 进行指令的自动发出和应答, 可以减轻训练负担和人员成本, 提升训练成效。

语音识别技术的关键在于语音引擎的设计。卡内基梅隆大学开发的 Sphinx 语音引擎是一个优秀的开源系统, 它对于英语语音的识别率较高, 在很多领域得到应用^[1]。与英语相比较, 由于汉语中的同音字和同义词比较多, 汉语连续语音识别

还存在不少问题, 特别是对于舰艇指挥训练系统的特定环境下的研究还鲜有报道^[2-3]。

根据舰艇指挥训练系统语音指令的非特定人、中等词汇量的连续汉语语音的特点, 本文选用了卡内基梅隆大学开发的 sphinx 开源语音识别引擎, 借鉴其在英语连续语音识别方面的特有技术, 设计并实现了一个用于舰艇指挥训练的专用语音识别系统。

1 Sphinx 语音识别引擎

Sphinx 是完全用 Java 语言写的先进的语音识别系统。它是通过卡内基梅隆大学 Sphinx 组、Sun 微系统实验室、三菱电器研究实验室和惠普等联合完成的, 其特点在于大词汇量、非特定人、连续英语语音的高识别率, Sphinx 系统的鲁棒性强、可扩充性很好, 并且代码开源代码, 其连续语音识别系统的基本流程如图 1 所示^[4]。

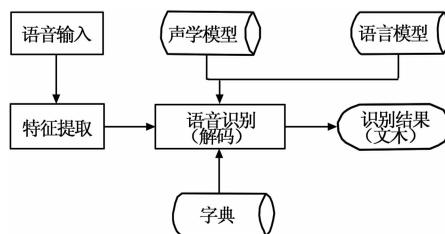


图 1 连续语音识别流程图

收稿日期: 2013-10-12; 修回日期: 2013-12-29。

作者简介: 景春进(1981-), 男, 江苏南通人, 硕士, 主要从事计算机通信技术方向的研究。

一个连续语音识别系统主要由语音特征提取、模型训练和语音识别等部分组成。连续语音识别系统的研究主要集中在声学模型和语言模型两个方面。本文主要采用 CMU 开发的工具箱中的 SphinxTrain, Cmuclmtk 和 Pocket Sphinx, 分别用来进行声学模型的训练、语言模型的训练和语音识别。

2 指令的语音识别关键问题

2.1 指令的语音特征分析

舰艇指挥调度指令是指舰艇在指挥调度系统中使用的通用规范用语, 一般由连续的汉字数字和简短的汉语调度指令组成, 该指令用语遵循一定的规范格式, 表 1 和表 2 给出了一组舰艇指挥调度语言的构成分析。

表 1 指令格式

被呼叫席位号	指令内容
021	请报告方位

表 2 指令回复格式

席位回复	回复内容
021 收到	当前方位是: 东经 118.36° 北纬 32.11°

由以上两个表的内容可以看出舰艇指挥调度指令具有一定的规律性, 主要由数字和舰艇指挥专用术语组成, 所用到的汉字大约为 400 个左右, 其样本空间数较小, 很适合语音识别技术的运用。

2.2 声学模型的训练

在连续汉语语音识别中, 声学模型的建立和训练是关键步骤之一, 通过声学建模, 可以估计待识别特征矢量序列所对应的语音识别单元, 从而完成特征矢量序列到语音识别单元的识别转换^[5]。声学模型的建立与语音发音密切相关, 这里我们结合汉语发音特点选择合适的声学训练单元进行模型的训练。在汉语连续语音识别中, 常用的单元包括: 词、音节、声韵母及音素等。汉语音节特点在于它由声母和韵母组成, 考虑到本文中的语音样本小, 且连续语音中存在协同发音的情况, 所以本文声学模型选择上下文相关的音素模型, 即比较常用的比较常用的三元音素模型, 以前述中的席位号“021”为例, 使用三元音素模型表示如下:

Silence Silence - Ling + er ling - er + yi er - yi + Silence (其中 Silence 表示静音)。本文中声学模型的训练采用的工具为 SphinxTrain。SphinxTrain 是 CMU 大学开发的开源声学模型训练器。此工具可以训练适用于 PocketSphinx 的半连续 HMM 声学模型, 也可以训练应用于 Sphinx 的连续 HMM 声学模型。SphinxTrain 工具中包含 4 个模块: Python 模块、头文件、库函数以及训练函数工具包。其声学模型训练的流程如图 2 所示。

利用 SphinxTrain 工具进行声学模型训练之前, 先通过 SphinxTrain 中的 wave2feat 文件对音频文件进行特征提取, 得到的 MFCC 特征参数文件, 并把它们作为声学模型训练的输入, 另外需要准备相关数据文件, 包括音素文件、主字典文件、补充字典文件、控制文件及脚本文件等。然后根据图 2 的流程图就可以进行训练了。最终得到 4 个声学模型参数文件:

均值文件、方差文件、状态转移矩阵和混合权重文件。

2.3 语音模型的训练

语言模型 (Language Model) 是描述自然语言内在规律的数学模型^[6]。构建语言模型的目的是建立一个能够描述给定词序列在语言中的出现概率的分布。在利用计算机进行语音识别时, 基于统计的语言模型被广泛使用。N-Gram 是大词汇连续语音识别中常用的一种语言模型, 汉语语言模型利用上下文中相邻词间的搭配信息, 在需要把连续无空格的拼音转换成文本时, 可以计算出具有最大概率的句子, 从而实现到汉字的自动转换, 无需用户手动选择, 避免了许多汉字对应同一发音的重码问题。本文中的指挥调度指令结构简单, 具有一定的规律性, 本文采用 Cmuclmtk 工具进行语言模型的训练, 通过统计大量文本数据得到以单个数字建立的 N-Gram 模型。因为 N 较大, 需要的存储空间和导致的运算时间会很大, 考虑到指挥调度指令中的汉字样本比较有限, 所以训练中主要采用 3-Gram 模型, 即某个词出现的概率仅依赖于前一个或者前两个词^[7]。例如以 10 个数字组成的汉语文本为例, 将若干数量的包含数字的语句作为输入, 通过 Cmuclmtk 工具箱处理后, 得到语言模型文件。

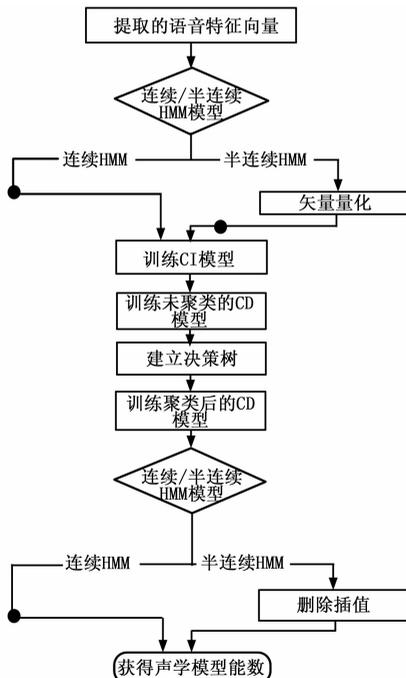


图 2 声学模型训练流程图

3 语音识别系统的设计与实现

本文的舰艇指挥训练系统中的语音识别子系统是基于 Sphinx 语音识别引擎而设计的, 其主要模块如图 3 所示。

指令语音通过麦克风录入后, 利用 MFCC 算法对每句话进行特征提取, 得到特征文件, 提供给声学模型进行训练。

语言模型是通过对所有指挥调度指令的文本进行训练后获得。

声学模型训练则对所有的指令语音进行训练, 包括数字汉语发音、非数字汉语发音及多音字发音 3 部分。

最后通过识别引擎的解码, 获得指令的文本, 并进行保存。

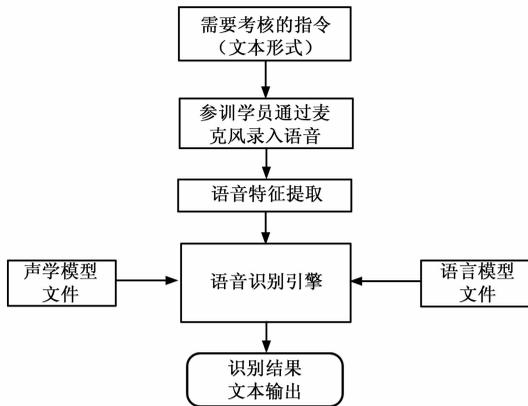


图 3 语音识别系统主要模块图

通过对最后识别出的语音指令的文本和考核指令的比较, 可以对训练结果进行分析, 给出训练成绩评定。

4 实验结果及分析

本文中语音识别系统的设计实现及实验环境均选择 Win7 下的虚拟 Linux 操作系统, jdk1.6, MyEclipse6.1, 以及语音识别工具 Sphinx 系列。选取的语音测试集为 100 条全数字语句和 368 条常用的舰艇指挥调度指令。参加测试的人员为男女各 3 人; 实验中每条指令识别正确与否的标准为: 数字识别全部正确且其他汉字识别超过 90%。实验的统计情况如表 3 和表 4 所示。

表 3 全数字语句识别情况

人员	数字语句数量	声学训练前识别数量	声学训练后识别数量	训练前后正确率(前/后)
男声 1	100	58	98	58%/98%
男声 2	100	60	96	60%/96%
男声 3	100	57	96	57%/96%
女声 1	100	66	98	66%/98%
女声 2	100	64	95	64%/95%
女声 3	100	62	98	62%/98%

上面两个测试集的实验结果表明, 系统在声学模型训练前, 由于测试人的语音发言、语速等因素的影响, 识别率比较低; 经过声学模型训练后, 识别率有了明显的提高, 尽管识别率没有达到 100%, 但是作为训练使用, 基本可以达到训练系统的要求。

表 4 指令语句识别情况

人员	数字语句数量	声学训练前识别数量	声学训练后识别数量	训练前后正确率(前/后)
男声 1	368	192	348	52%/94%
男声 2	368	196	346	53%/94%
男声 3	368	188	350	51%/95%
女声 1	368	195	351	53%/95%
女声 2	368	198	350	54%/95%
女声 3	368	201	353	55%/96%

实验中, 通过对未能正确识别出的情况做进一步分析, 发现主要是某些数字在发音时间间隔过短, 造成遗漏; 另外发言相同或很接近的某些词识别出错; 此外, 在语音间隔的时候, 偶尔有词被遗漏, 为此, 在软件设计中每句语音结束时, 通过输入键盘按键信息, 加入了结束标志, 从而避免上述间隔造成的影响。

5 结论

本文围绕舰艇指挥训练系统中语音识别技术进行了研究, 对舰艇指挥调度指令的语言特点进行了分析, 在此基础上借助 Sphinx 系统工具平台, 研究了连续汉语语音识别系统中特征提取方法, 声学模型的训练方法及语言模型的训练方法等, 进而构建了针对舰艇指挥训练系统的汉语连续语音识别系统。实验表明该系统基本达到了训练系统的要求。

参考文献:

[1] 马 萧. 汉语连续语音识别及其在移动机器人导航中的应用 [D]. 西安: 西安理工大学, 2007.

[2] 彭志勇, 王丹霞. 语音识别技术在 DRS 雷达模拟机系统中的实现 [J]. 中国民航飞行学院学报, 2006, 17 (2): 31-34.

[3] 张 健, 谭景信. 语音命令识别技术及其在雷达模拟机中的应用 [J]. 计算机工程与设计, 2010, 31 (3): 655-659.

[4] 徐 望. 连续语音识别的稳健性技术研究 [D]. 郑州: 中国人民解放军信息工程大学, 2006.

[5] 王晓龙, 关 毅. 计算机自然语言处理 [M]. 北京: 清华大学出版社, 2005: 18-22.

[6] 杨海燕, 景新幸, 曾招华. 基于 DSP 开发板的语音识别系统的研究 [J]. 计算机测量与控制, 2013, 21 (1): 210-211.

[7] 张仰森. 汉语统计语言模型训练样本容量的量化度量 [J]. 计算机科学, 2009, 36 (10): 222-224.

(上接第 2570 页)

参考文献:

[1] Horacio Diaz R, Ildefonso Harnisch V, Raul Sanhueza H, et al. Feeder Reconfiguration and capacitor placement in distribution systems: an approach for simultaneous solution using a genetic algorithm. [J]. ISSN. Revista Chilena de Ingenieria, 2010, 18 (1): 144-153.

[2] 徐 磊, 王耀南. 混合动力汽车电控网络总线技术研究 [J]. 计算机测量与控制, 2008, 16 (3): 321-323.

[3] 汪 森, 周国祥, 常安云. 基于 LIN 总线的车身控制系统设计 [J]. 合肥工业大学学报 (自然科学版), 2009, 32 (1): 94.

[4] 王 莉, 张 浩. 基于 CAN 总线的车身控制系统研究与应用 [J]. 组合机床与自动化加工技术, 2009, (10): 52.

[5] Dai C H, Chen W R, Zhu Y F, et al. Seeker optimization algorithm for optimal reactive power dispatch [J]. IEEE Transactions on Power Systems, 2009, 24 (3): 1218-1231.

[6] 吉武俊, 高 云. 集总式 BCM 车身电器控制模块电源管理系统的设计 [J]. 制造业自动化, 2013, 35 (4): 153.

[7] 蒋淑霞, 韩志刚, 周永军. 基于 CAN/LIN 总线车身网络优化设计与 CAN 通信实现 [J]. 仪表技术与传感器, 2011, (6): 48-50.