

陆空通话标准用语(英语)的语音指令识别技术研究

刘万凤¹ 胡 军^{1,2} 袁伟伟¹

(南京航空航天大学计算机科学与技术学院 南京 210016)¹

(南京大学计算机软件新技术国家重点实验室 南京 210093)²

摘 要 陆空通话标准用语(英语)的训练是空管模拟训练中的重要内容。对空管模拟训练中的语音指令自动识别问题进行了分析研究,包括:陆空通话标准用语基本特征的分析、语言模型的文法描述、指令特殊发音的识别处理、识别后处理方法以及声学模型训练方法等,并基于 Sphinx-4 语音识别引擎,设计实现了一个语音指令识别系统 AIRS (ATC Instruction Recognizer System)。系统实验数据分析表明,声学模型训练后的语音识别正确率可以达到空管模拟训练的需求。

关键词 语音识别, ATC 指令识别, Sphinx-4, 陆空通话

中图分类号 TP31 文献标识码 A

Research on Technology of Voice Instruction Recognition for Air Traffic Control Communication

LIU Wan-feng¹ HU Jun^{1,2} YUAN Wei-wei¹

(College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)¹

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)²

Abstract Training of air traffic control(ATC) communication in English is the main content of ATC simulator training. This paper focused on the voice instruction recognition in ATC simulator training which includes the analysis of the basic characteristics of ATC communication, the grammar description for language model, the recognition processing in special pronunciation of instruction, the processing method after recognition and the way of acoustic model adapting. Based on the excellent speech recognition engine Sphinx-4, we designed and implemented a voice instruction recognition system, AIRS (ATC Instruction Recognition System). The experimental results show that the accuracy rate of voice instruction recognition after the acoustic model adapting can reach the demand for ATC simulator training.

Keywords Speech recognition, ATC instruction recognition, Sphinx-4, Air traffic control communication

随着我国民航运输事业的快速发展,空中交通流量急剧增长,对培养高水平的空中交通管制(Air Traffic Control, ATC)人员的需求越来越大^[1,2]。目前,对空中交通管制学员的培训主要通过空管模拟训练机完成。陆空通话标准用语(英语)^[3]的训练是空管模拟训练中的一个非常重要的内容。

语音识别^[4,5]是通过计算机及软件将语音信号自动转变为相应的文本或命令的过程。20世纪60年代,英国实现了世界上第一个计算机语音识别系统。现在,已存在一些企业和实验室(如 Nuance, IBM, 微软, CMU(卡耐基梅隆大学)等)研究的语音识别引擎^[6,7]在通用办公应用场景中有较好的语音识别效果,但是,目前将语音识别应用于空中交通管制领域的相关研究工作比较少。

本文对陆空通话标准用语(英语)(又称为 ATC 指令)的语音识别问题进行了分析与研究,并基于开源语音识别引擎 Sphinx-4^[8]设计实现了一个可用于空管模拟训练的语音识别软件系统 AIRS(ATC Instruction Recognizer System)。具体内容如下:第1节给出了空管模拟训练和语音识别引擎

Sphinx-4 的概述;第2节对 ATC 指令语音识别过程中的一些关键问题进行了分析并给出解决方法,包括 ATC 指令的基本特征分析、语言模型的设计、特殊发音的识别、识别后处理和声学模型训练等;第3节给出了语音识别系统 AIRS 的设计架构和系统各功能模块的说明;第4节则给出了系统实验及数据分析;最后是相关研究工作的比较分析,以及对本文工作的意义以及进一步工作的简要讨论。

1 空管模拟训练与语音识别

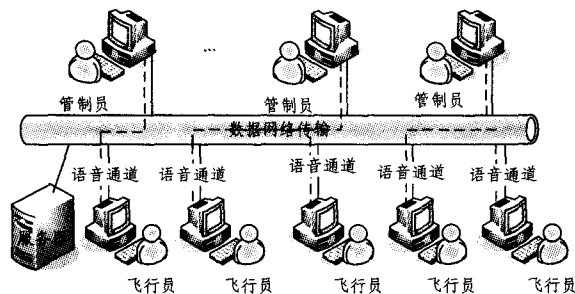
1.1 空管模拟训练

空管模拟训练机是用来模拟真实的空中交通管制过程的仿真系统^[1],是训练管制学员和飞行学员的重要方法和有效措施。目前的空管模拟训练机架构都设有管制员席位和飞行员席位,如图1(a)所示。在培训管制学员的过程中,由教员在虚拟机长席位依据管制学员发送的语音指令,通过手动的输入,完成相应航空器的操控,并通过专用的语音通道人工地给出应答信息。由于教练员人数有限,通常是管制学员轮流

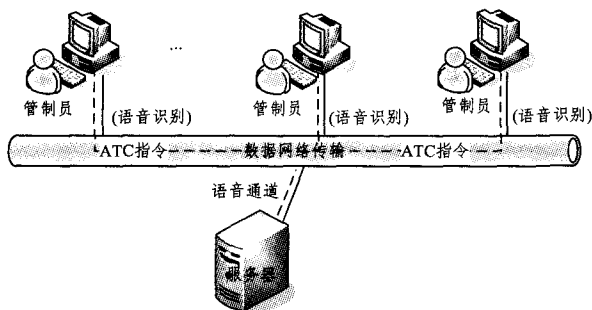
到稿日期:2012-09-26 返修日期:2012-12-30

刘万凤(1988—),女,硕士生,CCF 学生会员,主要研究方向为软件工程,E-mail:liuwanfeng2010@126.com;胡 军(1973—),男,副教授,CCF 高级会员,主要研究方向为软件分析与验证、嵌入式系统、形式化方法;袁伟伟(1987—),男,硕士生,主要研究方向为软件工程。

担当多名飞行员配合其他学员培训。这样,难以有效保证管制学员的训练时间和训练效果,计算系统设备也存在较大的不必要冗余。



(a) 典型的模拟机席位设置图



(b) 采用语音识别技术后的模拟机席位设置图

图1 语音识别使用前后模拟机席位设置图

管制人员和飞行人员的通信用语称为陆空通信用语。统计资料显示,在飞行事故和不正常情况下,管制人员和飞行人员之间的通信用语错误造成的因素占65%^[9]。因此,陆空通信用语的规范、统一、无歧义对飞机飞行的安全有着非常重要的意义,对陆空通信用语标准用语进行训练也是模拟训练的重要内容。中国民航局参照国际民用航空组织的ATC用语手册制定了《空中交通无线电通信用语标准(MH/T 4014—2003)》^[3](以下文中简称《标准》),其语法结构规范、无歧义、样本空间数小且工作环境噪声低,比较适合采用语音识别技术。

图1(b)为采用本文所研究设计的陆空通信用语识别系统,空管模拟训练机的席位设置构架。此时,由各管制员席位的计算机完成ATC语音指令的识别及指令解析,并通过数据网络传送到后台服务器处理,这为实现具备语音识别与应答功能的自动飞行员席位奠定了技术基础。

1.2 语音识别引擎 Sphinx-4

Sphinx-4^[8,10,11]是CMU采用java语言开发实现的先进语音识别引擎,是较早实现非特定人、大词汇量和连续语音等识别技术的系统。20世纪80年代,CMU的J. K. Baker以及IBM的F. Jelinek等人将隐马尔科夫模型(Hidden Markov Model,简称HMM)^[12,13]应用于语音识别,取得了重大的突破。HMM也因此成为语音识别研究的主导方法,其显著地改善了语音识别引擎Sphinx-4的识别性能^[11,14],使得Sphinx-4已成为当前最好的识别引擎之一。目前,Sphinx-4的语音识别分为3种:基于语音命令的识别、基于文法的连续语音识别和基于听写的连续语音识别。

图2所示为Sphinx-4的识别原理框架,其主要包括3个模块:FrontEnd, Linguist和Decoder,各个模块的功能如下。

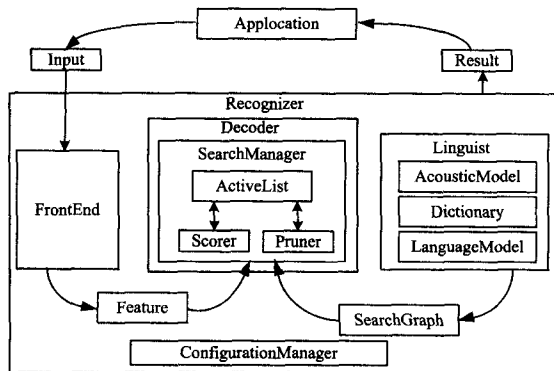


图2 Sphinx-4 识别原理框架图

1)FrontEnd:由一个或多个处理计算信号的并行处理器模块构成,主要根据输入的语音信息提取其中的一个或多个语音信号,将其参数化为一系列的特征序列,用于Decoder模块的输入;

2)Linguist:结合语言模型中描述的文法、发音字典中发音以及声学模型的特征参数,得到Decoder模块使用的搜索路径;

3)Decoder:将FrontEnd模块得到的特征序列与Linguist模块产生的搜索路径进行匹配完成解码过程,产生最终的识别结果并输出。

本文研究中采用CMU的Sphinx-4语音识别引擎,主要考虑以下几点:

(1)Sphinx-4是开源的语音识别引擎,源代码获取方便,便于功能的增加与修正;其用Java语言开发,可移植性好,可以在多个平台上运行。

(2)Sphinx-4识别正确率高,经过声学模型训练后的识别率可以达到90%以上,为语音识别技术在空管模拟训练的实用化奠定了基础。

(3)Sphinx-4可以自定义语言模型,并识别特定文法所描述的语言模型。陆空通信用语标准用语中指令简明、无歧义、准确,不仅便于文法描述,而且可以避免识别出非指令用语,这非常有利于提高识别正确率。

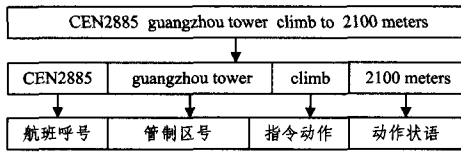
2 ATC指令语音识别中的关键问题

ATC指令是在涉及飞行安全的空中交通管制中使用,其特定的应用场景带来了一些语音识别中的特殊问题。本节中首先分析了ATC指令的基本特征并给出指令格式的设计,然后给出了语言模型描述文法的设计,接下来对ATC指令中特殊发音的识别进行了分析,然后说明了识别后处理的方法,最后研究了声学模型训练的问题。

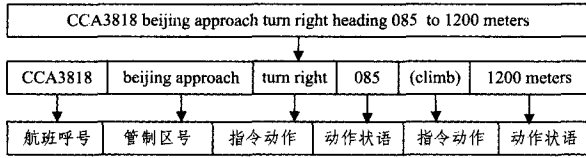
2.1 ATC指令的基本特征分析

ATC指令是管制人员和飞行员之间的通信用语规范,是以英语为基础但并非普通英语的通信用语,其发音别具一格,基本特征是发音高度清晰、单词的单一意向性及结构的祈使性^[2,15]。本文中所研究的ATC指令是《标准》中所规定的标准ATC用语。

ATC指令的应用场景中,涉及到飞机的起飞、进近、着落、航线飞行、地面滑行等多个飞行阶段,指令数量庞大、种类较多。本文对《标准》中的ATC指令进行分析后,发现空中交通管制人员和飞行人员之间的典型通信用语遵循一定的格式,图3(a)、图3(b)给出了两条典型的ATC指令结构的分析。



(a) 简单类指令的结构(如:高度变化)



(b) 复合类指令的结构(如:调整航向并改变高度)

图3 典型ATC指令结构的分析

因此,本文工作中将一条完整的ATC指令格式设计为:
“航班呼号”+“管制区号”+“指令动作”+“动作状语”

其中,航班呼号表明接收ATC指令的航空器编号,一般是由航空公司呼号和航班编号组成。航空公司呼号按照国际民航组织指定的无线电呼号给出,如air china(CCA),china eastern(CES)等等,通常为字符串形式。航班编号是由4位数字任意组成的字符串,如2338,4677等等。

管制区号是ATC指令的发送者,一般由管制单位名和管制类别(均为字符串)组成。管制单位名指明管制单位所在地,如北京(beijing)、广州(guangzhou)、成都(chengdu)等等。管制类别指明管制指令的所属范围,如塔台(tower)、进近(approach)、地面(ground)、区域(control)等等。

指令动作和动作状语则用来表明具体的指令,一条指令中可以只包含一条指令动作和动作状语,如图3(a)简单指令中的climb和2100 meters,也可以包含多条指令动作和动作状语,如图3(b)复合指令中的turn right和085、climb和1200 meters。在某些情况下,航班呼号、指令动作等部分还可以缺省,这些缺省的信息需要在语音识别过程中进行识别和补全,如图3(b)中的climb。

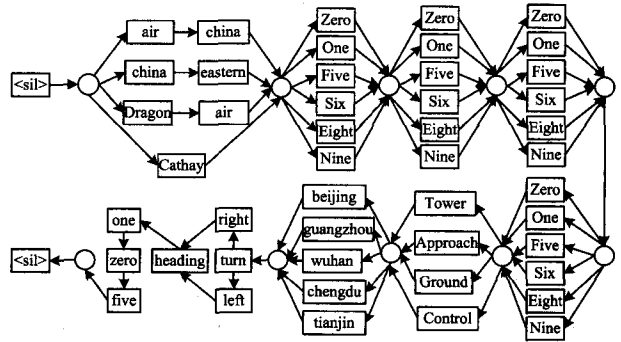
2.2 语言模型设计

语言模型是描述自然语言内在规律的数学模型^[16]。构造语言模型是语音识别必不可少的一部分。在语音识别引擎Sphinx-4中,语言模型提供单词级的语言结构,与声学模型和发音字典共同配合,产生语音识别的匹配搜索路径^[11]。目前,语言模型有两种主要的描述方式:基于文法的语言模型描述,基于统计语言模型^[17]的描述。前者主要通过人工方式按照某种文法规则编写,适用于语言结构规整、有规律可循的语言,后者借助概率参数来评估字母、单词、句子的分布概率,主要应用于基于听写的连续文本数据的处理。考虑到ATC指令是飞行员和管制人员之间的通信用语,要求发音清晰、指令明确、语言简练,且满足上一节中有规则的结构特征,本文中采用了基于文法的语言模型来进行ATC指令的语言描述。

语音识别引擎Sphinx-4描述基于文法的语言模型所使用的格式是JSJG(Java™ Speech API Grammar Format)^[18,19],该格式依据BNF范式定义语法规则,具有平台独立并与供应商无关的特性^[11],是Sun公司为应用语音识别技术专门提供的接口规范。图4(a)中示例为所设计的航向改变时ATC指令的核心JSJG文法描述。

```
Public <HeadingChangeIn>=<(<Flight>)(<AirControl>)(<HeadingChange>);
<Flight>=<(air china|china eastern|Dragon air|Cathay)(<Digit>)(<Digit>)(<Digit>);
<AirControl>=<(<ControlPlace>)(<ControlType>);
<ControlType>=<Tower|Approach|Ground|Control>;
<ControlPlace>=<beijing|guangzhou|wuhan|chengdu|tianjin>;
<HeadingChange>=<turn (right|left) heading one zero five>;
<Digit>=<Zero|One|Five|Six|Eight|Nine>;
```

(a) 航向改变的ATC指令的文法描述



(b) 文法识别过程中转化的非确定有限自动机

图4 文法描述和非确定有限自动机

语音识别引擎启动时将自动加载定义的语言模型文法,将其转化为非确定有限自动机(NFA),并依据该自动机决定可识别的单词或语句。图4(b)是语音识别引擎Sphinx-4加载图4(a)中JSJG文法过程转化的非确定有限自动机。此外,在《标准》中所规定的ATC指令包含飞机飞行的多个阶段,每个阶段又包含多种指令,指令种类复杂、指令数量庞大。若语言模型只用一个文法文件描述,则此文法文件的内容将非常庞大,语音识别引擎加载过程以及语音识别的速度将会受到影响。考虑到管制学员在单次训练中其实只涉及某一个特定的飞行阶段,本文中按飞行阶段的不同对ATC指令集进行了分类,设计多个文法文件,使得可以针对飞行的不同阶段加载不同的文法文件,提高了语音识别的效率。

```
grammar ApproachIn;
public <ApproachIn>=<(<Flight>)(<AirControl>)(<Approach>)(<Airwaitins>)(<Expectapptime>);
<Flight>=<(<FlightType>)(<Digit>)(<Digit>)(<Digit>)(<Digit>);
<FlightType>=<air china|china eastern|china northern|china southern
|hainan|china southwest|Dragon air|Cathay|shenzhen air>;
<AirControl>=<(<ControlPlace>)(<ControlType>);
<ControlType>=<Tower|Approach|Ground|Control>;
<ControlPlace>=<beijing|guangzhou|wuhan|chengdu|tianjin>;
<Approach>=<(<Approach_1>)(<Approach_2>)(<Approach_3>);
<Approach_1>=<(cleared|proceed) via ((<Designation>)(<DetailsofRoute>));
<Approach_2>=<cleared visual approach runway (<RaceNumber>);
<Approach_3>=<cleared visual approach runway (<RaceNumber>) Followed by
circling to runway (<RaceNumber>);
<Airwaitins>=<((cleared|proceed) to (<Navigation>)) ((maintain|climb to|descend to)
(<Height>)) hold ((<Heading>)) as published expected
approach [further] clearance at (<Time>));
<Expectapptime>=<(<Expectapptime_1>)(<Expectapptime_2>)(<Expectapptime_3>);
<Expectapptime_1>=<no delay expected>; <Expectapptime_2>=<expected approach (<Time>)>;
<Expectapptime_3>=<revised expected approach (<Time>)>;
<Heading>=<(zero|one|two|three)(<Digit>)(<Digit>);
<Digit>=<Zero|One|Two|Three|Four|Five|Six|Seven|Eight|Nine>;
<Height>=<((seven|eight|nine|six) hundred)(((<Digit_thousand>)((one|two|three))
thousand [(<Digit_hundred>) hundred]);
<Digit_thousand>=<One|Two|Three|Four|Five|Six|Seven|Eight|Nine|ten|eleven>;
<Digit_hundred>=<One|Two|Three|Four|Five|Six|Seven|Eight|Nine>;
<Time>=<(<TimeDigit_1>)(<Digit>)(<TimeDigit_2>)(<Digit>);
<TimeDigit_1>=<zero|one|two>; <TimeDigit_2>=<zero|one|two|three|four|five>;
<Designation>=<(<Letter>)(<Letter>)(<Letter>)(<Letter>);
<DetailsofRoute>=<(<Letter>)(<Letter>)((<Letter>)(<Letter>)<Letter>);
<Navigation>=<(<Letter>)(<Letter>)((<Letter>)(<Letter>)(<Letter>)(<Letter>));
<Letter>=<Alpha|Bravo|Charlie|Delta|Echo|Foxtrot|Golf|Hotel|India|Juliett|Kilo|Lima|Mike
|November|Oscar|Papa|Quebec|Romeo|Sierra|Tango|Uniform|Victor|Whiskey|Xray|Yankee|Zulu>;
<RaceNumber>=<(zero|one|two|three)(<Digit>)((<Multi_Runway>));
<Multi_Runway>=<Right|Left|Center>;
```

图5 进近管制ATC指令的JSJG文法描述示例

具体而言,假定一架飞机从地点A飞到地点B,整个飞行过程中ATC指令用语可以分为塔台管制用语、进近管制用语和区域管制用语。塔台管制用语中包含了飞机准备从停机坪推出到飞机起飞后或者飞机准备降落到完全停在停机坪过

程中的通信用语;进近管制用语包含飞机起飞后离场到航路飞行以及飞机从空中航路飞行到进场过程中的通信用语;区域管制用语包含巡航过程中的通信用语。依据这3类指令,本文设计了3个不同的文法文件描述语言模型。进近管制用语的核心JSGF文法描述如图5所示,塔台管制用语和区域管制用语限于篇幅,这里略过。

2.3 特殊发音的识别

由于在陆空通话中对ATC指令理解的正确性要求非常高,因此,常用英语中发音相近的字母或单词,如B和D、G和J等必须能在陆空通话中清楚地区分开来,尤其是指令中的航班号、航向、航线代号、导航台名称等由数字或字母组成字符串的发音,必须保证不能在管制人员和飞行人员之间造成理解的偏差和混淆。在《标准》中依据国际民用航空组织的规定,制定了一系列的字母、数字、专用名词的特殊发音及其发音规则的要求,如A读成Alpha,9读成niner等。

在本文的工作中,使用语音识别引擎Sphinx-4的发音字典所提供的39个基本音素来构造这些特殊的发音,即:字母、单词和特殊发音之间的映射可以用音素的组合排列(即发音序列)来实现,使得语音识别引擎可以有效地识别特殊发音的字母、数字单词及其组合。Sphinx-4所提供的音素集如表1所列。

表1 音素分类表

音素种类	音素
元音音素	AA,AE,AH,AO,AW,AY,EH,ER,EY, IH,IY,OW,OY,UH,UW
爆破音音素	B,D,G,K,P,T
摩擦音音素	CH,JH
送气音音素	HH
摩擦音音素	DH,F,S,SH,TH,V,Z,ZH
流音音素	L,R
鼻音音素	M,N,NG
半元音音素	W,Y

以下给出陆空通话标准用语中主要的3类特殊发音及其发音序列的构造。

表2 数字发音及其发音序列

数字	英文	英语读法	发音序列
1	ONE	WUN	W AH N
2	TWO	TOO	T UW
3	THREE	TREE	T RI Y
4	FOUR	FOW-er	F OW AH R
5	FIVE	FIFE	F AY F
8	EIGHT	AIT	EY T
9	NINE	NIN-er	N AY N ER
			T AW S AH N D
1000	THOUSAND	TOU-SAND	T AW Z AH N D
			T AW Z AH N

(1)数字特殊发音的识别处理:ATC指令格式中航班号和动作状语部分都是由多位数字组合而成,且这些数字在指令发布时要求按照《标准》逐位进行发音,如123在通话过程中会被读成one two three,而不是常见的“一百二十三”。此外,在涉及飞行高度数字、航向数字的指令时,其发音规则也存在与日常发音规则较大的不同,如当航向为25时,仍然按3位数025来发音,即:zero two five。因此,为了使得语音识别引擎能够有效地识别这些与日常发音方式及其发音序列都大有不同的数字及组合,我们对《标准》中所有类别指令中的数字发音方式进行了研究,设计了8个核心的非常用发音的

数字发音序列,如表2所列,加上Sphinx-4发音字典中已存在的zero,six,seven,hundred等4个单词及其发音序列,总共12个核心的数字发音序列,可以用于在不同语义场景下(如:飞行高度层、航向、时间、无线频率、速度等)的发音识别。

(2)字母特殊发音的识别处理:ATC指令中航线代号、导航台名称、跑道号等都是由多位字母(A-Z)或字母/数字共同组成,且多位字母在指令发送或接收过程中都是逐位读出的(如航线代号AWT03的发音为Alpha Whiskey Tango zero three)。如前所述,在《标准》中,为避免发音混淆,对26个英文字母给出了新的发音规则,即分别用ALPHA-ZULU等单词的发音来替换A-Z字母的发音,如航线代号BARC的发音分别为Bravo Alpha Romeo Charlie。

通过对Sphinx-4的分析,发现可以在自带的发音字典中找到ALPHA-ZULU中的24个单词及其发音序列。因此,本文对字母特殊发音的处理方法是:首先将不存在的2个字母(FOXTROT、JULIETT)的发音序列添加到发音字典中,然后,在后期识别后的处理中将语音引擎识别出的ALPHA-ZULU按照ATC指令规则替换成字母A-Z。表3所列为字母特殊发音的部分示例。

表3 字母发音及其发音序列

字母	英文	英语读法	发音序列
F	FOXTROT	FOKS TROT	F AA K S T R AA T
J	JULIETT	JEW LEE ETT	JH UW L IY EH T
A	ALPHA	ALPHA	AE L F AH
B	BRAVO	BRAH VOH	B R AA V OW
...	...	...	...

(3)专用名词发音的识别处理:陆空通话标准中以及具体的管制指令往往还涉及到空管专用名词,如:ILS(仪表着陆系统)的发音为I(AY) L(EH L) S(EH S),另外还有一些航空公司名、地名等这些名词的发音序列是发音字典中没有的,也需要将这部分的单词汇集整理,并将发音序列添加到发音字典中。

2.4 识别后处理

由于语音识别引擎所识别的语音输出结果只是一串由字母单词组成的字符串文本,不包含任何的结构与语义信息,无法直接作为模拟机服务器端处理所需的格式,因此,本文工作中对识别引擎输出的字符串设计了一个处理过程(见图6)。其基本思想为,以2.1节分析得出的ATC指令特征为基础,设计了一系列的语义结构的分析与替换规则,并使用Javacc(Java Compiler Compiler)^[20,21]来自动构造基于这些规则的语法/词法分析器。

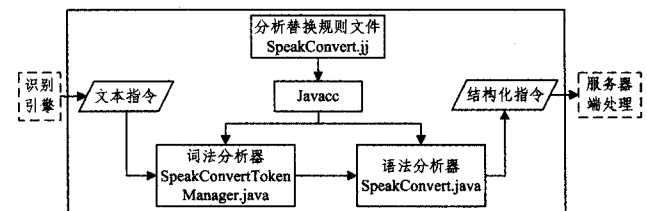


图6 Javacc 替换处理流程图

图6中的分析与替换规则文件(SpeakConvert.jj)是我们基于2.1节中的指令结构信息所设计的各类规则,包括:航班呼号的替换规则;航向、速度、时间、高度、频率等数字的替换规则;航线代号、导航台名称、跑道号等字母或字母/数字共同组合的替换规则等。图7给出了分析与替换规则文件中的核

```
void Start(String&InLoc buffer) {
    Token t;

    FlightMap(InLoc(buffer)) //飞机航公司名称和航班号替换的动作处理
    //清空动作A的动作表，主策略航向、高度、速度的替换。一次清空A
    "vz=00000" << InLoc.append("vz=10000"); //对替换量表的动作处理
    "vz=00000" << InLoc.append("vz=10000"); //对时间的替换动作处理
    NumberMapping(InLoc(buffer)) PointMapping Number2Fz(InLoc(buffer)) //导航点替换策略的动作处理
    NumberMapping(InLoc(buffer)) //航向替换策略的动作处理
    NumberMapping(InLoc(buffer)) //高度替换策略的动作处理
    HeightMapping(InLoc(buffer)) //航向、速度替换的动作处理
    LetterMap(InLoc(buffer)) //航班号替换的动作处理
    LetterMap(InLoc(buffer)) Number2Fz(InLoc(buffer)) //字母或数字和数字替换的动作处理
}

<C>
```

通过 Javacc 就可以分别自动构造能识别纯文本字符串中 ATC 指令相应语义单位的词法分析器和语法分析器,并进行替换生成符合 2.1 节中形式的结构化指令。这样处理的另一个好处是:当模拟机服务器端所需处理的指令格式发生变化时,只需要修改相应的分析与替换规则,即可自动生成相应的解析器。

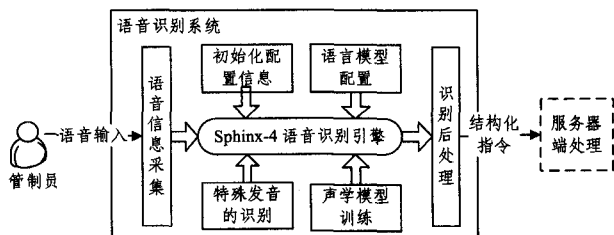
声学模型^[22,23]提供语音单元到模型之间的映射, 将其将输入的语音单元和上下文情景按模型图进行检索, 并与语言模型和发音字典共同配合, 构造语音识别的匹配搜索路径^[11]。Sphinx-4 采用隐马尔科夫模型 (HMM) 对声学特征进行建模, 可以实现针对非特定人、大词汇量的连续语音识别。但目前 Sphinx-4 中使用的是基于“华尔街日报”的典型美式发音的声学模型, 与国内管制学员的英语发音特征存在较大不同, 直接应用时, 效果并不理想。

```

graph TD
    Start([开始]) --> UserLogin[用户登录]
    UserLogin --> DisplayContent[/训练内容显示/]
    DisplayContent --> LoadRecord[/训练记录加载/]
    LoadRecord --> DecisionContinue{继续上次训练?}
    DecisionContinue -- 否 --> ResetRecord[训练记录重置]
    ResetRecord --> DecisionContinue
    DecisionContinue -- 是 --> MicCollect[麦克风语音采集]
    MicCollect --> CreateFile[创建语音资料库文件]
    CreateFile --> DecisionReRecord{当前训练重录?}
    DecisionReRecord -- 是 --> DecisionContinue
    DecisionReRecord -- 否 --> DecisionContinueTrain{继续语音训练?}
    DecisionContinueTrain -- 否 --> ExtractFeature[声学特征文件提取]
    ExtractFeature --> GetInfo[获取统计信息]
    GetInfo --> UpdateModel[更新声学模型]
    UpdateModel --> End([结束])
    DecisionContinueTrain -- 是 --> DecisionContinue
  
```

具体而言,管制学员在初次使用时登陆训练系统,系统会自动显示训练的内容并加载上次训练的记录,管制学员依据系统的提示决定是否继续上次的训练,之后通过麦克风录入具有个人特色的语音信息,录用过程中,学员可以选择当前训练的重新录音、暂停当前录音及结束训练。此时,训练过程会依据麦克风采集的语音信息提取声学特征文件,获取统计信息并最终更新声学模型,从而完成整个训练过程。

Sphinx-4,设计并实现了涵盖起飞、进近、着陆、航路飞行、地面滑行等多个不同空中交通管制场景的语音识别系统 AIRS,其系统框架如图9所示。主要模块包括:语音信息采集,初始化配置信息,语言模型配置,特殊发音的识别,声学模型训练及识别后处理等。下面给出各个功能模块的说明。



- 语音信息的采集:其功能是监听麦克风的输入,采集管制学员的语音信息,将语音信息按特定的采样率、单/多声道等属性进行处理并保存在缓冲区作为语音识别引擎的识别输入,目前的设置是 16kHz,单声道。

- **语言模型配置:**主要负责设计与配置不同飞行阶段的语言模型文法,使得系统可以根据不同的空管交通场景来配置选择适合的语言模型供语音识别引擎加载,完成不同阶段的训练。

• **声学模型训练:**负责对各类指令进行训练,以提高识别的效率,语音训练的界面设计如图 10 所示。考虑空管领域中对数字、字母等发音具有较高的语音识别要求,将训练分成 3 部分:

b)数字训练,专门针对航班号、高度、速度、航向、时间中数字部分的语音训练;

Figure 1 is a schematic diagram of a 1000-MVA power transformer with 160000 kVA capacity. The diagram illustrates the internal structure and cooling systems of the transformer. It shows three main windings: HV (High Voltage), MV (Medium Voltage), and LV (Low Voltage). The HV winding is connected to a 1000 MVA busbar, the MV winding to a 100 MVA busbar, and the LV winding to a 10 MVA busbar. The cooling systems are labeled as 'HV cooling system', 'MV cooling system', and 'LV cooling system'. The diagram also shows the 'HV cooling system' and 'MV cooling system' connected to a 'HV cooling system' and 'MV cooling system' respectively. The LV cooling system is connected to a 'LV cooling system'.

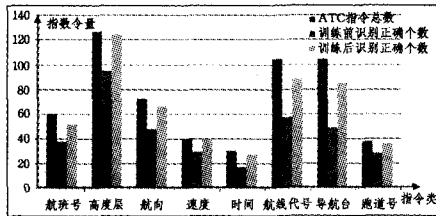
- 识别后处理:此模块主要是将识别出的字符串形式的指令进行解析重组,即将信息中的航空公司名、航班号、航向、导航台名称、跑道号等具有特定含义的单词替换成相应的结构化数据形式,以方便模拟机进一步对指令信息进行处理。

4 实验及分析

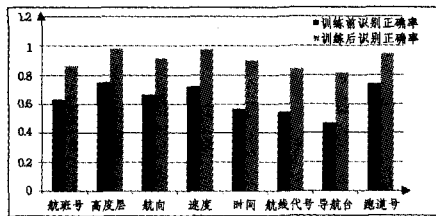
语音识别正确率是判断语音识别结果好坏的重要标准,也是判断语音识别系统优劣的一个重要指标。在本节中给出了标准的 ATC 指令在 AIRS 系统中识别正确率方面的实验及统计分析。AIRS 系统的设计开发与实验运行环境是:Java SDK 1.6, Eclipse SDK 3.4.0, Windows 7 操作系统平台,双核处理器等。本文依据《标准》,设计了 574 条 ATC 指令的实验数据。

每一条 ATC 指令都至少包含着航班号、航向、速度、高度、时间、航线代号、导航台名称、跑道号等方面的改变,这部分都是由多位数字或字母组成的。考虑到空管领域中对各类数字、字母的识别要求非常高,在以下的实验统计分析中,采用了最严格的标准,即除非几位数字和字母同时识别正确,否则整个 ATC 指令的识别都将视为不正确。本文的实验结果数据及分析均以此为基础。

此外,由于训练后的声学模型能够更好地适应个人语音特征,提高语音识别效率,因此,本文的实验中也给出了声学模型训练前、后两组不同的实验对比。图 11 和图 12 分别给出了甲、乙两名管制学员声学模型训练前后数字类(包括航班呼号、高度层、航向、速度、时间)和字母类(包括航线代号、导航台名称、跑道号)ATC 指令识别正确数和正确率的统计。

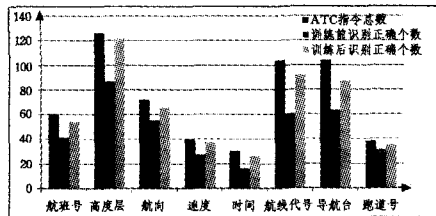


(a) 甲声学模型训练前后正确数

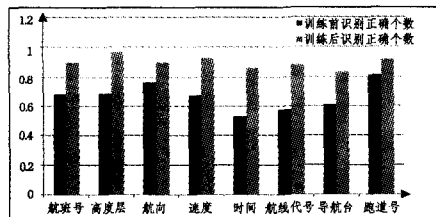


(b) 甲声学模型训练前后正确率

图 11 甲声学模型训练前后对比图



(a) 乙声学模型训练前后正确数



(b) 乙声学模型训练前后正确率

图 12 乙声学模型训练前后对比图

图 11、图 12 实验结果显示, AIRS 系统在声学模型训练前,由于管制学员的发音语速、声调等特征与 Sphinx-4 原有的标准美式发音的声学模型存在较多的差异,识别正确率并不理想(如图 11(b)和图 12(b)中识别率基本在 80%以下)。

但是,经过声学模型训练后的 ATC 指令的识别率有较大幅度的提高,如图 11(b)中航班号的识别率由 63.3%提高到 86.7%,图 12(b)中跑道号的识别率由 81.5%提高到了 92.1%,而且,数字类指令的识别效果明显比字母类的识别效果要好。虽然识别率没有达到 100%,但这样的效果已经可以达到空管模拟训练机对语音识别正确率的基本要求,为 AIRS 系统可以嵌入到空管模拟训练机提供了可能。对实验中那些判定为没有识别出的指令进行进一步分析,我们发现 90%左右的情况是因为指令中仅存在一位或两位数字未识别成功,如 ATC 指令 CCA 8352 beijing approach increase speed to 300 knots 识别出的结果为 CCA 8355 beijing approach increase speed to 300 knots。也就是说,这条所谓未识别指令中其实绝大部分都是正确信息。我们可以借助这些正确信息来设计与模拟机训练内容相关的匹配方法,在未来工作中进一步提高识别率。

5 相关工作

目前,已经有不少针对通用办公、日常生活领域中的语音识别产品,如股票查询系统、机器人接线员、铁路自动票务中心等,其核心的语音识别引擎包括 Nuance 公司的 Naturally Speaking^[24]、微软的 Speech SDK 5.1^[25]、IBM 公司的 ViaVoice^[1]等等,但是将语音识别技术应用到空管领域的研究工作还比较少。文献[2,15]中较早地给出了基于 ViaVoice 的一种空管模拟训练机的语音应用系统的设计方案,但缺乏核心问题的分析,也没有具体的实现过程,语音识别文法文件的设计也存在局限性,不能涵盖陆空通话标准用语中常用的指令。文献[26]主要只是介绍语音识别技术语音识别文法的设计规则、雷达管制人员席位模块及机长席位模块的概要设计,缺乏语音训练模块的分析和设计,没有实验数据,语音识别结果正确率也没有进行相应的统计和分析。文献[27]中主要探讨了在雷达模拟机中应用语音命令识别技术的可能,文中强调了语音命令识别技术、语音识别算法、语音信号的输入输出过程,也给出了雷达模拟机中自动机长席位实现过程中的各个子模块的说明,但缺乏系统性的说明,实验中涉及的指令也不完整。和上面相关工作相比,本文的研究理清了空管语音识别中的关键问题,对系统的整个处理框架给出了详细的说明,并给出了 AIRS 系统的设计与实现。此外,我们融合了声学模型的训练,有效地提高了识别正确率并给出了声学模型训练前后识别正确率的实验统计分析,在语音训练过程中,还对识别易错的数字和字母进行了专门的训练,有效地提高了识别结果的正确性。

结束语 本文对语音识别技术在空管领域中的应用进行了研究,分析了语音识别过程中空管指令的特征分析、语言模型的设计、特殊发音的识别、识别后处理方法及声学模型训练方法等关键问题。在此基础上,以 Sphinx-4 为核心设计并实现了一个 AIRS 空管语音识别系统,并详细地给出了相关 ATC 指令实验结果的统计和分析。实验数据表明,训练后的

AIRS 系统的识别正确率可以满足空管模拟训练机的基本训练需求。本文工作为进一步实现具备自动机长席位的空管模拟训练机奠定了技术基础。

进一步的工作包括以下几个方面:

1) 陆空通话标准用语指令的扩充。本文工作中设计并实现的指令主要包括塔台管制用语、进近管制用语和区域管制用语中常用的指令,部分不常用的管制用语尚未添加(如紧急情况处理等),下一步可以将这部分指令也添加到语音识别子系统中,以使整个系统更加完善。

2) 语音合成技术的使用。本文的 AIRS 系统很好地为嵌入空管模拟训练机提供了可能,接下来可以采用语音合成技术,给出 AIRS 系统识别的 ATC 指令的一次或多次应答信息,实现自动飞行员席位。

3) 容错机制的设计。目前,初次使用系统的识别正确率并不能达到 100%,下一步可以根据已经识别的绝大部分正确信息设计相关的指令匹配方法,提高识别正确率,使 AIRS 系统能够具有一定的容错性。

参考文献

- [1] 由扬,徐肖豪.空管模拟训练机的语音用系统设计[D].南京:南京航空航天大学,2002
- [2] 聂党民,张林昌.塔台模拟器中语音识别子系统的开发与应用[D].北京:北京航空航天大学,2003
- [3] 李京东,杨新渥,毛笑歌,等.空中交通无线电通话用语[S].MH/T 4014-2003.北京:中国民用航空总,2003
- [4] Hunt M, Yates J, Bridle J. Automatic speaker recognition for use over communication channels[C]//Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 77). 1977, 2: 764-767
- [5] Al-Qatab B A Q, Aïnou R N. Arabic speech recognition using Hidden Markov Model Toolkit (HTK)[C]//International Symposium in Information Technology. 2008: 557-562
- [6] Maheswari N U, Kabilan A P, Venkatesh R. Speaker independent speech recognition system based on phoneme identification [C]//Proceedings of International Conference on Computing, Communication and Networking (ICCCN 2008). St. Thomas, VI, 2008: 1-6
- [7] Sharp R D, Bocchieri E, Castillo C, et al. The Watson speech recognition engine[C]//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Munich, 1997: 4065-4068
- [8] Lee K F, Hon H W, Reddy R. An Overview of the SPHINX Speech Recognition System[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1990, 38(1): 35-45
- [9] Zemrowski K M. Impacts of Increasing Reliance on Automation in Air Traffic Control Systems[C]//Proceedings of the 2nd Annual IEEE Systems Conference. Montreal, Que, 2008: 1-6, 7-10
- [10] Tsontzos G, Orglmeister R. CMU Sphinx4 speech recognizer in a Service-oriented Computing style[C]//Proceedings of the 2011 IEEE International Conference on Service-Oriented Computing and Applications (SOCA). Irvine CA, 2011: 1-4
- [11] Walker W, Lamere P, Kwok P, et al. Sphinx-4: A flexible open source framework for speech recognition[R]. TR-2004-139. Sun Microsystems Laboratories, November 2004
- [12] Rabiner L R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition [J]. Proceedings of the IEEE, 1989, 77(2): 257-285
- [13] Abushariah A A M, Gunawan T S, Khalifa O O, et al. English digits speech recognition system based on Hidden Markov Models[C]//Proceedings of International Conference on Computer and Communication Engineering (ICCCE 10). Kuala Lumpur, 2010: 1-5
- [14] Baker J K. The Dragon system-an overview[J]. IEEE Transactions on Acoustic, Speech and Signal Processing, 1975, 23(1): 24-29
- [15] 由扬,徐肖豪.空管模拟机的 IBM ViaVoice 技术实现研究[J].中国民航学院学报,2002,20(3): 6-9, 15
- [16] 王晓龙,关毅,等.计算机自然语言处理[M].北京:清华大学出版社,2005
- [17] Keyes L, O'Sullivan A, Winstanley A. Graphical object recognition using statistical language models[C]//Eighth International Conference on Document Analysis and Recognition (ICDAR 05). 2005, 2: 1095-1099
- [18] <http://java.sun.com/products/java-media/speech/forDevelopers/JSRG/>
- [19] Holland W, May D, Baca J, et al. A Unified Language Model Architecture for Web-based Speech Recognition Grammars[C]//2006 IEEE International Symposium on Signal Processing and Information Technology. 2006: 294-299
- [20] <http://javacc.java.net/>
- [21] Kodaganallur V. Incorporating language processing into Java applications: a JavaCC tutorial[J]. Software, IEEE, 2004, 21(4): 70-77
- [22] Huang X D, Lee K F, Hon H W, et al. Improved Acoustic Modeling for the SPHINX Speech Recognition System[C]//Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 91). Toronto, Ont, Canada, 1991: 345-348
- [23] Schultz T, Waibel A. Language independent and language adaptive acoustic modeling for speech recognition[J]. Speech Communication, 2001, 35: 31-51
- [24] <http://naturallyspeaking.net/>
- [25] <http://www.microsoft.com/china/community/program/original-articles/techdoc/Cnspeech.msp>
- [26] 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
- [27] 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