

语音翻译系统技术分析

The Technical Analysis of Speech Translation System

宗成庆 陈肇雄 黄河燕

(中国科学院计算技术研究所机译中心 北京 100080)

摘 要 The speech translation is considered to be the grand challenge for modern computer science and engineering. In this paper, some typical speech translation systems are introduced, and some key problems in speech translation system design are discussed in detail. By presentation of some suggestions, the authors hope to have a good help to speech translation research.

关键词 Speech translation, Speech recognition, Machine translation, Speech synthesis

语音翻译(Speech Translation)技术作为一门综合性的计算机应用技术,近年来得到了广泛的关注。有关专家曾指出,语音翻译是自然语言处理、语音识别及其人工智能研究的最终目标之一,是当今世界对计算机科学和工程最大的挑战^[1],其应用效果如何在某种程度上代表了当今世界计算机技术发展和应用的水平。

语音翻译的研究涉及领域广泛,不仅涉及语音识别及合成,自然语言处理和理解、机器翻译以及人工智能等许多高难度理论问题,而且涉及电子学、声学、心理学等许多相关学科,因此,迄今为止,语音翻译技术除在少数严格限定的应用领域和规定的条件下取得阶段性应用成果(诸如:限定领域的自动电话翻译系统,导游信息服务系统,饭店服务系统,售票服务系统以及交通服务、列车时刻表查询、电话号码查询等语音信息服务系统)外,许多理论问题和技术难题仍有待于进一步研究和探索。

一、国内外研究现状

尽管与语音翻译相关的语音识别与合成、机器翻译等技术的研究已有一定的历史,然而真正的语音翻译研究及技术开发只不过是近十几年的事情。一般地认为,语音翻译的研究起始于八十年代^[1]。十几年来,包括美国的卡内基-梅隆大学(CMU)、SRI、瑞典的 Telia 网络电讯公司和日本的 ATR (Advanced Telecommunications Research, Interpreting Telephony Research Lab.) 以及德国的联邦政府科学研究与技术部等在内的许多著名大学和科研机构都投入了大量的人力和资金从事这一尖端技术的研究和开发工作,并在一些应用领域取得了重要的阶

段性成果。

1.1 国外语音翻译研究

CMU 机器翻译中心于 1989 年开发了国际上第一个语音翻译系统—Speech Trans 日-英语音翻译系统,该系统能够将日语会话翻译成英语会话,其限定领域是医生与病人之间的会话。Speech Trans 系统是语音翻译技术的里程碑^[1],自诞生以来,以美国 CMU 和日本 ATR 为代表的一些大学和研究机构先后开发了一批语音翻译系统的实验系统和原型,见下表^{[1][2]}所示。

表 1 国外语音翻译系统

系统名称	开发时间	开发单位	翻译语种	应用领域	词汇量
Speech Trans	1989	(美)CMU	日-英	医生与病人	
SL-TRANS	1989	(日)ATR	日-英	会议登录	1335
Phi Dialog	1989	(美)CMU	日-英	会议登录	
MINDS	1989	(美)CMU	英	数据库查询	1000
JANUS	1991	(美)CMU	英-日、德等	会议登录	
SLT	1993 开始	SRI, Telia	英-瑞、法	航班查询	1381

在我们所见到的资料中,只有 JANUS 系统报道了系统翻译的正确率(78.2%)。

目前,欧美等发达国家正联合日本、韩国等亚洲国家协作开展包括机器翻译、语音翻译和对话翻译等多国语言的自然语言处理系统的研究^[2]。

1.2 国内语音翻译现状

我国的语音翻译研究起步较晚,1990 年左右四川大学开发了国内第一个也是迄今为止国内报道的唯一的—一个英汉语音翻译系统^[3],其限定主题为航

空订票及信息查询,依赖于特定讲话人,识别词汇量为150个,句式21种。另外,北方交大、四川大学、中科院自动化所、声学所等单位先后开发了应用于铁路售票、交通旅游信息查询等领域的人机语音接口系统^[5,6]。相对而言,我国的语音翻译技术研究无论从规模上,还是从实现技术和研究单位的数量上与国外相比还有一定的差距,国内大部分工作都停留在单项技术的研究上,如语音识别的研究,语音合成的研究,机器翻译的研究和语音人机接口等相关内容。因此,如何针对汉语语音的处理特点及其汉语与相应外语之间的翻译特点,研究与汉语相关的自然语言语音翻译系统是一个十分引人注目的研究课题。

综上所述,目前国内外的语音翻译研究大多停留在原型开发或实验性系统研究阶段,其应用领域都受到严格限制和约束,系统识别的词汇量一般在1000单词左右,而且都是从一种语言到另一种语言的单向语音翻译,无论其应用领域的广泛性,语音识别和翻译的正确率,还是系统本身的灵活性和健壮性等诸方面,离最终实用的目标还有相当的距离。

二、语音翻译技术分析

2.1 基本工作原理

一个从源语言到目标语言的单向语音翻译系统通常情况下由语音识别、机器翻译和语音合成三大模块构成(图1)。

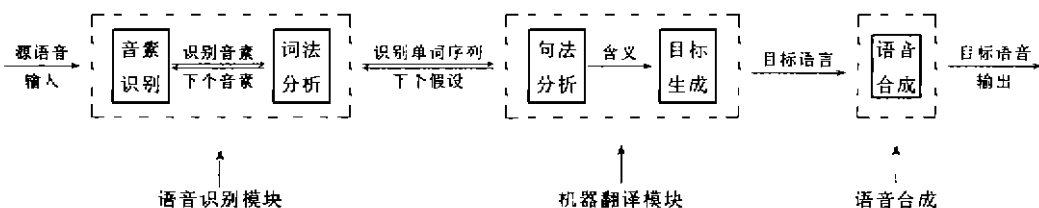


图1 语音翻译系统基本原理

语音识别模块是整个语音翻译系统的基础,识别精度如何,直接影响到整个系统的译准率。要成为一个实用的语音翻译系统,语音识别模块必须具有高精度、非特定讲话人、大词汇量、连续语音的实时识别能力。机器翻译模块是语音翻译系统的重要组成部分。它接受来自语音识别模块的结果,通过对讲话者意图的理解和分析,结合语音专家知识和会话常识,对其进行必要的处理,然后,利用语法规则和语义规则进行语法分析,形成语法分析树,最终生成目标语言。语音合成模块将机译模块生成的目标语言转换成相应的语音信号输出。语音识别、机器翻译和语音合成三大模块有机结合为一体,相应模块之间应具有良好的系统接口。实际上,这三大模块在很大程度上是从完成的功能上划分的,在实际应用系统的实现中,它们之间常常紧密结合,在许多系统中,是将部分语法分析任务与词法分析工作进行一体化处理。

另外,在整个语音翻译系统中,包括词汇知识,语法、语义规则,发音规则等在内的知识库系统也是必不可少的。双向对话语音翻译系统是两个互逆的单向语音翻译的有机结合。

2.2 面临的困难

语音翻译系统由语音识别、机器翻译和语音合成三大模块构成,因此开发语音翻译系统的主要困难也就分别来自这三大技术模块^[7~10]。

1. 关于语音识别。四十年来,语音识别和理解的研究离人们在不同应用领域中完全实用化的要求还有一定的距离,其困难在于语音信号的复杂性和可变性:

(1)说话人自身的可变性:同一个人在不同的环境、心理状态下,或采用不同的说话方式,其语音信息会有很大差异。即使上述条件相同,一个人的语音还要受长时间变化的影响。另外疾病等因素也会引起一个人的发声系统和语音的改变。

(2)不同说话人之间的可变性:由于说话人的社会背景、方言以及声道形状和长度的不同,造成了不同人在同一词上的发音有很大差别。

(3)语音的模糊性:语音的模糊性既有说话人的原因,也有语音本身的原因,包括说话人的因素、近音词、语音环境的影响等。相同的音节、词汇随语音环境和在语流中的位置不同,其语音特性会产生很大差异,音量、语调、重音、发音速度和连读方式等都有可能发生改变。

(4)噪声的干扰:噪声的干扰能明显地妨碍语音

特征的准确提取和辨识,使系统性能大幅度下降。因此,要想使语音识别走向实用化,就必须有效地解决噪声问题。

以上各种因素的影响,给语音识别研究带来了极大的困难,要想彻底解决这些问题,就必须结合数字信号处理、人工智能、模式识别、机器翻译等各学科的技术特点。目前,一方面应着重于语音信号特征参数的研究,力求反映语音信号的生理和听觉特性。另一方面应加强识别模型和方法的研究,提高正确率和鲁棒性。此外,对于连续语音识别还应充分利用连续发音的规律和构词、句法、语义等多方面的知识。

汉语是世界上使用人数最多的语种。普通话是汉语的标准语音,具有结构简单的特点:汉语以单音节为发音单位,每个汉字一个音节,由单韵母或声母、韵母相拼而成;汉语的音系比较简单;汉语是有调语言,具有阳平、阴平、上声、去声、轻声 5 种声调;在汉语音节结构中,韵母段能量大,时间长,具有准周期性,动态谱包络相对稳定。

汉语的上述特点使之比较适合于实现语音识别。目前我国的汉语语音识别研究已经从特定人、有限词汇、孤立音节输入发展到非特定人、大词汇量、连续语音输入,从技术上已从单一的语音模型发展到相当规模和性能良好的语音模型^[9]。但是,在实现技术上,除了上述提到的系统性能和鲁棒性等诸多方面仍有待于深入研究外,如何对语音识别的结果(拼音流)进行有效的后处理,得到确切的汉字形式表达,也是汉语语音识别中有待于进一步研究的关键性问题之一。

2. 关于机器翻译。文本到文本的机器翻译研究已取得十分可观的成就,并已推出了若干商品化的系统,但许多技术问题仍有待于进一步研究和探索,翻译的正确率和系统性能都有待于进一步提高,而且语音翻译系统中的机器翻译与一般的文本到文本的机器翻译相比,有许多重要的区别。主要体现在如下几个方面:

(1)语音翻译中较多地出现非规范句子,翻译机制必须具备处理各种不规范句子的能力。口语会话中,不可能象书面语一样,说话者的每个语句基本都符合句法和表达习惯,省略也是常见的现象,而且,同一讲话内容可能多次重复。另外,由于语音识别模块对噪声等干扰因素处理不当也可能产生一些多余的成份或非法成份。因此,翻译机制必须具备处理由此而引起的各种不规范句子的能力。

(2)分析机制必须具有理解讲话者意图的能力。在日常会话中,同一个词的词意往往依赖于周围词汇、说话者的年龄、性别、说话风格以及特定环境等变化而各有差异,不同的句子之间的含义可能是相关的,也可能毫无关系,而且不象文本中有标点注释,而且口语中还经常使用一些无法与文字对应的声音表示特定的含义。因此,翻译机制必须能够理解或基本理解讲话者的意图,否则,无法对原话进行正确地翻译。

(3)翻译机制与语音识别模块之间应有良好的模块接口。语音识别模块往往提供几种可能的识别结果,并将最有可能的假设首先送给翻译模块,因此,翻译机制常常需要与语音识别模块之间进行信息交流和反馈,设计友好的模块接口就显得非常重要。

(4)在文本到文本的翻译系统中,系统可以借助译前编辑和译后编辑等人工手段来提高系统的译准率,而在理想的实时性语音翻译系统中,一般不应有人工干预的要求,即使有也应尽量减少人工干预的时间,把更复杂的任务交给机器去完成,这就给翻译机制提出了更高的要求。

总之,语音翻译系统中的机译机制与文本翻译中的机译机制相比更复杂,更需要相关技术的支持,实现难度也更大。

3. 关于语音合成。语音分析与合成技术也是自然语言处理领域中活跃的分支之一。国内关于语言合成的研究开始于七十年代末,经过众多科学家的艰苦努力,已经取得了可喜的成果,尤其在清晰度方面已达到相当高的水平。但目前国内的语音合成系统仍然需要深入地研究^[10]:(1)加强韵律特征研究,提高语音合成的自然度;(2)增强文本处理能力。

综上所述,语音翻译系统是综合多种技术的集成系统,开发成功的语音翻译系统需要包括语言学家和计算机专家及信号处理专家等在内的多方面人员共同协调合作。除了上述每一部分提到的几个方面以外,整个系统如何达到实时性要求也是系统实现必须解决的问题之一。

三、关于语音翻译系统设计的讨论

由于语音翻译系统具有上述各种复杂特性,因此给系统设计带来较大困难。笔者认为,除了遵循一般系统设计的基本原理之外,设计语音翻译系统应充分考虑如下几点:

(1)设计好的知识表示模型。知识表示问题一直

是人工智能领域的一个重要问题。在语音翻译系统中,如何设计有效的知识表示模型来描述各种语言学知识、语音知识、会话常识和翻译规律等也至关重要。这种表达模型不仅应该能够准确地刻画推理机制(包括语音识别中的推理机制和机器翻译中的推理机制)所需要的各种知识和常识,而且必须有利于系统的机器学习机制进行知识获取和知识扩充。

(2)建立完备的专家知识库。专家知识库不仅包括语音知识、语言学知识和各种有关的常识,而且应提供如何运用这些知识的方法和经验,并且知识库的搜索和知识运用应该是高效的。

(3)建立良好的文法体系。无论是语音识别中用于音素识别的文法体系,还是机器翻译中用于语言分析和转换的文法体系,都应该是形式化较好的方便实用的文法描写体系。一方面,这些文法体系必须便于语言工作者描述和不断修改各种音素规则、词法规则和语法规则;另一方面,又必须便于高效分析机制的构造和运行。

(4)多种自然语言处理模型相结合^[11,12]。在自然语言处理中,无论是基于规则的推理方法还是基于语料库的统计学方法都有其自身的弊端,无论在语音识别中,还是在机器翻译中,任何一种单一的分析方法都难以胜任对如此复杂问题的推理和分析,因此,象 SL-TRANS 系统和 Φ_M DIALOG 系统那样,借鉴各种方法的优点,建立大规模真实语料库,采用基于规则的推理方法和统计学方法相结合的分析策略,彼此取长补短,可能是唯一正确的出路。

(5)系统应具有自学习能力。系统在运行使用过程中,应该具有自我完善、自我调节的能力,包括词汇量的扩充、异常情况处理、经验的积累以及常识积累等。

一个实用的语音翻译系统开发是一个庞大的工程项目,需要多方面人员的统一协作,难度大,开发周期长。因此,鉴于目前国内的技术水平和研究现状,笔者提出如下系统开发模式:

1. 着眼于受限语言的研究,开发受限的语音翻译系统。尽管语音翻译研究的最终目标是实现通用的语音翻译,但在目前各种技术条件还不够完善的情况下,着眼于受限语言的研究,开发受限的语音翻译系统将是一条可行的途径,并且受限的语音翻译

系统仍然具有极大的应用价值和科学意义。这里所说的“受限”指应用领域受限,词汇量受限,会话句型受限和发音方式(接近普通话的清晰发音)受限等。

2. 集中力量,联合开发。目前国内大量的研究机构都在从事语音识别、机器翻译和语音合成等单项技术的研究,许多工作有重复,不利于充分发挥各自的技术优势,因此,集中力量,互相补充,联合开发,才是一条容易成功的科研道路。

结束语 语音翻译技术的研究是一个极具挑战性的课题。随着语音识别和机器翻译等技术的不断提高,语音翻译的研究必将取得可喜成果。尽管国内对这项技术的研究起步较晚或刚刚起步,但是中文语音识别及合成、与中文相关的机器翻译技术的研究已有一定基础,因此,与中文相关的语音翻译技术研究条件已基本成熟,我们希望在不久的将来会开发出较为实用的外汉或汉外语音翻译系统。

参考文献

- [1] Hiroaki Kitano, Speech-to-Speech Translation, a Massively Parallel Memory-based Approach, Kluwer Academic Publishers, 1994
- [2] SRI Cambridge Report, Spoken Language Translation: First-Year Report, Jan. 1994
- [3] Research Actives for C-STAR ■ Members, 1996
- [4] 扬家沅、林道发等,连续英汉语音翻译系统的设计和实现,声学学报,17(5)
- [5] 俞一彪、袁保宗,连续语音识别中句法结构知识的利用,中文信息学报,7(1)
- [6] 申凌、毛育等,汉语语音理解系统,全国第四届人机通讯学术会议论文集,1996
- [7] E. Klein, F. Veltman, Natural Language and Speech, Symp. Proc. Brussels, Spring-Verlag, Nov. 1991
- [8] Yorick Wilks, Theoretical Issues in Natural Language Processing, Lawrence Erlbaum Associates, 1989
- [9] 王仁华,语音识别测试报告,计算机世界报,1996,3.25,评测专刊
- [10] 张家录,汉语语音性能评测结果,计算机世界报,1996,3.25,评测专刊
- [11] 周明、黄昌宁等,统计与规则并举的汉语句法分析技术模型,计算机研究与发展,31(2)1994
- [12] 王挺、陈火旺、史晓东,语料库和机器翻译,计算机科学,1996,2