

研究表明,对于功率受限系统(如扩频通讯系统),图4情形较有利;对于能量受限系统(如脉冲压缩雷达等),图5情形较有利<sup>[4]</sup>。

## 2. 应用

SAW 卷积器在通讯领域中的应用主要有两方面,一是作为可程序匹配滤波器,多用于扩频系统,二是作为高速运算器件,在合成孔径雷达中获得应用<sup>[5]</sup>。下面给出一些扩频系统中的具体应用。

(1) 用于快速同步电路中,如图6所示,图6(a)是一种逐次逼近法。假如电路中的各元件均为理想器件,则当接收信号同步头中第N组码到来时,两信号的同步误差将 $\leq T/2^{N+1}$ 。该方案电路简单,可用于要求不太高的场合。图6(b)是一种一次性捕获方案,它利用双向计数器进行延时控制,从而缩短了捕获时间,但相对来讲成本较高。

(2) 用于解调和扩,如图7。用两个分别为全“0”和全“1”的本地扩频码与接收信号进

行相关检测,再将检出的相关峰送比较器判别,即可同时达到解调和扩的目的。

(3) 此外,它还可以用于各种直接序列扩频接收机及一些其它系统,参见<sup>[6-7]</sup>。

## 四、结 束 语

本文对 SAW 弹性卷积器的原理及应用进行了简要的介绍,并着重讨论了快捕原理及其应用。事实上,它远不止这些用途。目前,影响 SAW 卷积器应用的主要方面是卷积器的卷积输出效率和价格,但是它的优良的信号处理功能也具有很强的吸引力,所以,随着通讯和 SAW 技术的进一步发展,SAW 卷积器必将获得更广泛的应用。

本文错误之处在所难免,欢迎指正。

在本文的整理过程中,得到李忠诚研究员和竹志年、仰书耀副教授的指导和帮助,特此表示感谢。

## 参 考 文 献

- [1] 冯所椿等,电子学报,2(1983),73—77.
- [2] 武以立等,声表面波原理及其在电子技术中的应用,国防工业出版社,1983,277—285.
- [3] Lewis M. F., et al., IEE Proceedings, 1984,131A (4), 204—215.
- [4] John C. H.,IEEE Ultra. Symp., 1984, 121—126.
- [5] Gautier H., AFCS, 1982, 428—441.
- [6] Chatterjee Amitava, et al., IEEE Trans., SU-32 (1985), 745—759.
- [7] Glisic Savo G., 电讯技术(译文),27—4 (1987),75—79.

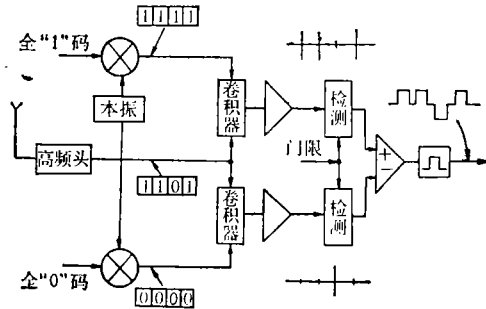


图7 SAW 卷积器用于解调和扩

# 国外声音理解的研究动向

陈阿林 张素

(西南师范大学数学系计算机软件专业)

一九八七年七月三十日

本文介绍了世界各国目前声音理解研究动向。主要介绍了美国 DARPA 第二次声音理解研究计划和日本国际电气通信基础技术研究所(AIR)的自动翻译电话系统计划。包括各项基础研究,实验系统结构,各处理部分所用方式、功能、以及计划达到目标等。并对我国声音识别,特别是汉语声音识别,

研究提出相应的建议。

## 一、前言

声音是人类传递信息的手段之一。进入信息化社会的今天,作为人工智能的重要组成部分,声音的合成、识别、理解的研究工作,近年来非常活跃,取得了许多成果。特别是进入知识工程阶段后,使用人工智能(AI)技术,声音识别和理解的研究水平有了很大的提高。本文拟介绍美日等国在声音识别和理解方面的研究状况,同时就我国此方面的研究,提出一些建议。

## 二、国外声音理解研究动向概述

对声音的研究,起源很早,但只是在计算机技术的发展中,对声音的研究,才有了较大发展。50年代后期,人们开始进行语音识别研究,经过十几年的探索,得出了非常重要的结论:即语音不是一种单纯的声音信息,为识别语音,必须进行理解。并且了解到,人对声音的理解是一个复杂的智能活动过程。70年代初,美国的DARPA(Defence Advanced Research Projects Agency)开始了第一次声音理解系统研制计划。并研制成功由1000个单词组成的连续声音识别系统Harpy。另外,卡内基·梅隆大学(CMU)还研究开发Hearsay II系统,它是在“黑板”模型的AI系统概念下进行声音理解的系统,因硬件配置及声音知识不完善,该系统尚未实现。日本的各有关研究所、大学也完成了许多连续声音识别实验制系统<sup>[1]</sup>。NTT武藏野通研完成的系统是以新干线坐席预约为对象的声音理解系统,按预约项目的声音来切分发音,进行语音识别,并通过对话模型进行合成声音作为应答的装置。法国的许多研究中心,研究机关,大学都开展了从发音机制到理解系统的研究工作。农西计算机科学研究中心完成了查询天气消息的口呼语言理解系统原型,电讯研究中心完成了专人声控拨号模型系统。法

表1 各国声音理解系统的有关计划及机构

国家	研究机关或研究计划	研究目的
美国	DARPA(CMU, MIT, BBN, TI, etc)	大字表连续语音识别、理解系统。
日本	ICOT	第五代计算机技术。
日本	ATK 自动翻译电话研究所	自动翻译电话系统的各要素技术基础研究等。
美国	CMU(卡内基梅隆大学)	办公文书的大字表声音输入系统提问-应答系统,语法校正基础研究。
美国	Bell 研究所	语音识别会话理解基础研究。
美国	MIT(麻省理工学院)	英语及其它语种的声音合成,对话理解的基础研究。
美国	耶鲁大学	文章理解模型,文章理解存贮模型的基础研究。
美国	CSLI 研究中心	自然语言和计算机语言的统一综合模型。
美国	SRI	用人工智能(AI)的手段,进行自然语言处理的研究及与数据库的接口方面研究。
美国	BBN	语言含义的表现手段及对对话理解模型基础研究。
FC(欧洲)	埃斯普利计划	与信息通信技术有关的声音识别合成及语言处理研究。
英国	阿贝尔计划	新一代计算机研究和与其相关连的高性能声音识别,合成研究。
法国	GRICO计划	声音合成、识别、理解及声音DB建立,音素标注研究等。
法国	信息实验室	语言声学,知觉结构及合成-识别-对话理解,用于机器人。

国科研中心的信息实验室广泛进行语言的声学 and 知觉结构的研究,以获得语言信号的声学 and 语音学两个层次上的最佳知识,为更有效地理解声音信息创造有利条件。

进入80年代,声音识别、理解进入了新的

研究阶段。主要对基础和硬件设备环境进行研究，着重开发各类匹配基准的单词声音识别装置，同时建立了音素标注数据库，语言数据库，并以 AI 技术来完成声音理解。参加研究的人员大大增加，各种技术不断涌现。各国纷纷提出国家级别的研究计划，采用国际交流方式进行研究。

美国作为 DARPA 的计算机战略计划之一，以 CMU, MIT, BBN 为中心的第二阶段声音理解系统的研究计划<sup>[2]</sup>，于 1985 年开始。该计划研究之重点是融合语言处理技术，特征基元的识别方法，分布式系统技术，AI 技术等，来建立声音理解系统。并建立音素标注的声音数据库。在欧洲，英、法等国的新一代计算机计划，无不包含声音识别/理解的研究。日本国际电气通信基础技术研究所 (ATR) 建立了 ATR 自动翻译电话分所，1986 年也开始了包含声音理解的自动翻译电话系统的研制计划<sup>[3-4]</sup>。为建立此系统，日本人开始了各种基础研究，包括声音合成，声音识别，分析理解，机器翻译，系统结构技术，声音数据库，语言数据库等。此研究工作是有组织的，长期的，它的技术水平的提高是实现系统的必要条件。日本人认为，此项计划是在世界范围内第一次提出，它的实现会对世界有所贡献。

国外声音理解系统的有关计划及其研究机构详见表 1。

### 三、美国研究状况

美国主要以 CMU, MIT, BBN, IBM, Bell 研究所为中心，推进声音理解系统的研究。

#### 1. DARPA 第二次声音理解系统计划

1985 年 1 月，作为 DARPA 的战略计划的一环，以 CMU 为主开始制定了“不定人大字表连续声音识别”计划，预计 4 年完成。研究单位还包括 MIT, BBN, TI, NBS, SRI 等。预算投资为 1 千万美元/年。它为全美该研究领域中最大的计划，其目标是识别由 10000 个单词构成的各种不定人连续声音，要求处理的实时

性达到或接近人类正常对话速度。系统以各种知识源(KS)进行分散-综合处理。由包含有超级计算机组成的分布式系统支撑。主要研究机关的任务如表 2。下面简述以本计划为中心的 CMU 声音理解研究状况。

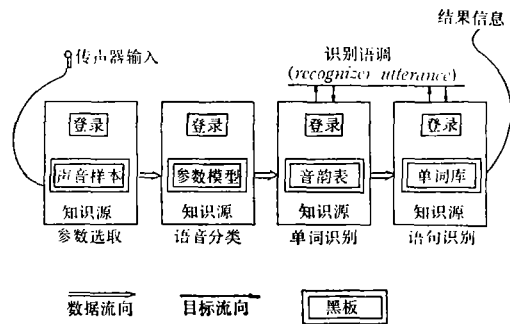


图 1 CMU 声音理解系统 (ANGELS)

表 2 DARPA 第二次声音理解系统计划研究项目分配情况

研究机构	分担项目
CMU (卡梅隆大学)	连续声音识别系统的结构，特征基元的音素识别算法；单词匹配定位识别算法；帧格文法处理句子及语法分析，语言理解算法。
TI (得克萨斯仪器公司)	收集供研究，评价系统的声音数据，专用硬件开发。
MIT (麻省理工学院)	对声音数据进行音素标注并协助 CMU 完成特征基元的音素识别算法。
BBN (Bolt Beranek and Newman Inc.)	用 HMM 模型的模式识别匹配方法，完成音素及单词的识别。
NBS (美国国家标准局)	声音数据库管理及判定系统评价标准、手段、方法。

CMU 目前研制的声音理解系统，受第一阶段声音理解系统之一的 Hearsay II 影响很大，系统结构如图 1 所示<sup>[5]</sup>。系统主要特征在于使用了知识源 KS (Knowledge Source) 进行工作。知识源 KS 是以共享的存储器作为“黑板”，KS 通过“黑板”接受信息知识。KS 设计为独立，并行地进行工作，并且以网络相互连接，在不同的计算机上加以实现。CMU 的该系统，由音韵识别、单词识别、语言处理、分布式系统

软件工具等构成。下面简单地说明各处理部分。

**音韵识别部分** 本部分用 16kHz 采样的语音声波作同步 DFT, 计算频谱信息, 然后对音韵序列各元素进行定位 (Locator), 找出其近似位置候选对象, 然后对其候选对象用分类的方法进行细微的音韵识别。此过程如图 2 所示。图中表出了四大类音韵元素的 KS。目前正在对异常音、音调结合等, 进行更为细微的定位和分类, 以得到更佳识别效果。此工作进一步是建立鼻音的分类方法。音韵识别结果是以经过定位, 分类产生的音韵表 (Lattice) 及相应概率一同送入单词识别部分。

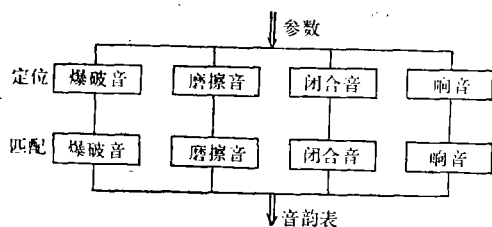


图 2 CMU 音韵识别部分

**单词识别部分** 本部分根据输入的音韵表, 进行单词的准确匹配定位。图 3 表示此处理部分。单词字典是考虑了异常音及结合语调的音韵序列, 以脱机方式建立。单词的定位匹配除用音韵表之外, 还参考音节、强度等信息, 以提高识别匹配正确率。作为结果是产生相应单词表送往语言处理部分。

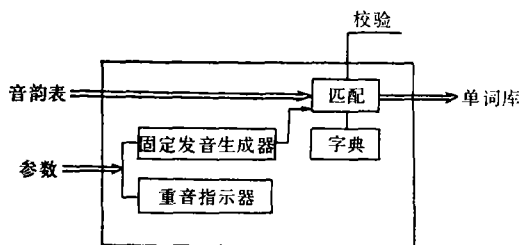


图 3 单词识别部分结构

**语言处理部分** 对输入的单词信息由帧格模型进行语句分析, 产生正确的文章。对音韵和单词识别, 均采用自底向上方式进行识别处

理的, 而在语言处理部分, 则是以自顶向下的处理方式进行, 该部分的结构如图 4 所示。为提高处理速度, 使用扩充 LR 文法分析, 产生单词序列候选对象, 同时参考单词产生的概率进行识别处理。此外, 本部分还将完成因音调组合或错误发音在前级未得出的发音的单词匹配工作。

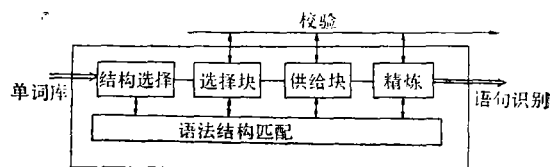


图 4 语言处理部分结构

**分布式系统** 它是实现声音识别系统的硬件, 软件支撑环境。硬件方面, 包括了各种计算机的网络连接、谱分析硬件实现等。软件方面, 主要研制了称为“Agora”的软件工具, 能使知识源模块化, 并通过“黑板”使各模块之间的控制及信息存贮方便进行。借助此工具, 希望达到 KS 的并行工作, 对全部知识的宏数据操作控制。本部分系统工具的研制, 将使用得克萨斯仪器公司 (TI) 收录的大量声音数据, 同时还使用 MIT 和 CMU 建立的音素标注声音数据库。

## 2. 其他研究机关

TBN 注特逊研究所的声音识别课题组研究的声音处理系统是以办公室书信文件为对象, 研制声音输入文字处理器。在语句处理级使用双重马尔柯夫模型、在音韵和单词处理级使用发音的 HMM (Hidden Markov Model) 模型来进行识别研究。

Bell 研究所的声音识别用改进 LPC 频谱匹配基元的声音识别算法, 进行数字语音识别, 同时还致力于接近实用水平系统的研制, 也对 HMM 识别匹配方式进行深入的研究。在语言级基础研究方面, 着重于常用对话语句结构分析。

## 四、日本 ATR 自动翻译电话研究所的计划<sup>[6]</sup>

该研究所声音识别课题,进行声音识别的基础研究及完成自动翻译电话系统的原型系统中声音识别子系统。其目标以四年为期,研制出由 1000 个单词构成的日文句子、文章的声音识别实验系统。在此基础上,用 3 年时间,将单词数量扩充到 4000,建立区分各文节发音的日语和英语会话声音的识别实验系统。本计划还注重建立加注音素标记的声音数据库及识别研究所用的标准单词开发研究。

声音识别系统根据“目标指向”的概念来建立其结构。各处理部分用模块方式联系,它们均可支持、使用共享分布式“黑板”。系统结构如图 5 所示。

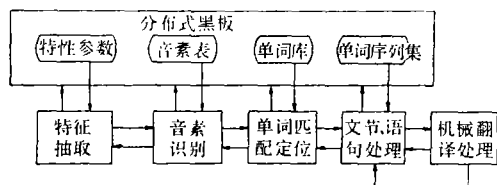


图 5 声音识别处理系统结构示意图

该系统用并行处理方式进行识别处理。它首先以自底向上方式处理初始数据流,产生出识别处理初步结果,然后再按自顶向下的处理方式,参考相应信息,进行更为精确的识别处理。控制方式采用目标指向,各处理部分可参考形成“作用器”(actor)的 KS 来进行工作,各作用器可共享分布式黑板并通过各分布式黑板完成其它作用器及 KS 的信息传授。所谓目标指向的概念,是指由并行处理,自底向上处理,自顶向下处理及“黑板”作为数据驱动,不断整理,综合“作用器”的起动,完成识别匹配工作。以下简述各处理部分。

**特征抽取** 以谱信息及 LPC 分析信息为基本数据,计算得出音素特征参数及共振峰轨迹的动态数据等。除频谱信息之外,语调频率,稳定程度、强度等韵律信息也作为辅助特征。本

应用声学

模块还可以由单词定位匹配等高层处理单元自顶向下进行启动。提取的特征参数,存入分布“黑板”特征参数区。

**音素识别** 本部分是频谱图记录信息的音素识别算法和 LPC 参数匹配音素识别算法组合构成。音素识别主要使用特征抽取得到的特征参数进行识别,同时还自顶向下频繁起动下级模块,即特征抽取部分工作,以便尽可能准确识别音素。本部分识别结果存入分布黑板的音素表中。

**单词匹配定位** 该部分是对音素表进行定位匹配,完成单词选取,当音素表信息不明确时,则通过音素识别模块,并参照特征抽取产生的特征参数再次起动特征模块,进行反馈性识别。在识别过程中,还灵活使用韵律信息,以辅助匹配定位。为保证结果的尽可能正确,对某些关键词的匹配将起动上级模块,即文节处理部分,以产生后续可能出现的字词语音预测信息。本部分处理结果存入分布黑板的单词库中。本部分承担着连续声音识别系统的中心处理任务。

**文节处理部分** 由单词匹配定位产生的结果,在本部分起动语法、语义及分布黑板中的韵律信息 KS,进行文节句子处理。本部分将预测后续可能接续单词信息,并且将结果送入前级单词匹配定位部分,同时还将使用前次预测信息及单词连接、与被连接关系的统计信息,产生可靠处理结果。为提高效率及精度,还使用 DP 匹配,树形搜索等技术手段。

## 五、各研究计划的特点和我国应采取的措施

除以上介绍外,英、法、欧洲各国对连续语音识别,理解的研究都很活跃。如 EC 的 CONST 209 计划<sup>[7]</sup>, ESPRIT 纲领<sup>[8]</sup>, 英国 ALVEY<sup>[9]</sup> 纲领; 法国 GRECO 及 PRC 计划<sup>[10]</sup>; NATO 的 RSG10 纲领<sup>[11]</sup>。它们的共同特点是: 均为国家级规模, 研究组织国际化, 充分重视国际间合作; 从研究角度和技术上, 认识到声音数据库的

• 45 •

重要性,并积极开始进行配备;同时开始制定声音处理系统的客观评价标准,主张确立客观评价手段、方法;研制综合多语种共同识别、合成;在处理中均利用 AI 手段,利用知识模块发展系统结构<sup>[12-13]</sup>。

在我国,从事声音合成、识别研究的单位不少,虽然开展了一些工作,进行了一些初步应用,但与国外相比,还相当落后,比较大且完善的识别、理解系统尚未见报道,对汉语的识别理解研究更少。研究力量分散,基础研究薄弱,人员少,资金不足,设备不完善,都不同程度阻碍我们对声音,特别是汉语声音识别的研究。最重要的是国家对此方面研究重视不够。为此建议:

1. 建立国家级别的研究计划,建立现代化科研中心,集中研究力量开展对汉语语音理解的研究,包括从基础研究到实验系统的建立等。

2. 重视广泛的国际交流,用派出去和请进来的各种方式,提高研究技术水平。广泛吸收国外此方面最新研究成果,经验,缩短我国的研究差距,和不必要的重复研究,促进超越汉语识别理解的国际水平。

3. 建立标准汉语发音的数据库,和汉语音节、音素(声母、韵母)、语调(四声)的标准。建立汉语声音处理系统的客观评价方法及标准。国家对人员、资金的安排从优考虑。

最后,从市场导向角度上看,声音识别,理解系统研制也将是大有可为的。美国某调查机

关预测,声音识别,理解设备的需求量,美国市场在1985年约为9200万美元,到1990年,将上升10倍。从我国计算机市场受国外产品冲击和占领的教训来看,加强此方面的研究,尽快形成我们自己的独特产品,也是有很大意义的。

## 参 考 文 献

- [1] 中田和男,日本音響学会誌,42(1986),920—923.
- [2] CMU Speech Group and Ready R., A Continuous Speech Recognition System, Vol. 1: Technical Proposal, Carnegie-Mellon University, 1984, 5.
- [3] 蒲口理一郎,角所收,音響学会誌,42(1986),396—402.
- [4] 廷野克彦,竹之内正一郎,蒲所收,情报处理学会第33回全国大会講論集,1986,6N-1—6N-3.
- [5] Adams D. A. and Bisiani R., Speech Tech. 4(1986), 14—23.
- [6] 鹿野清宏,樽松明,日本音響学会誌,42(1986),948—952.
- [7] Pirani G., Cost 209: Linking different European Language, Proc. Speech Tech., 1986, 208—211.
- [8] Roukens J., European Cooperative Research in Speech Technologies, Proc. Speech Tech, 1986, 195—197.
- [9] Holmes J. N., Speech Technology in the U. K-Alvey Program, Speech Tech. 1985, 44—47.
- [10] Mariani J., Speech Technologies in Western European——A Review, Part 1, Speech Tech. 1985, 48—57.
- [11] Moore R. M., The NATO Research Study Group on Speech Proc.: RSG10, Proc. Speech Tech. 1986, 201—203.
- [12] Mizoguchi R., Tsujino K. and Kakusho O., A Continuous Speech Recognition System Based on Knowledge Engineering Techniques, Proc. ICASSP86, 1986, 1221—1224.
- [13] 福田尚行,蒲口理一郎,角所收,信学技报,EA 84-6 (1985).

(上接第47页)

动圈传声器、压力式全指向性传声器、无变压器传声器、无线传声器、数字传声器等的近期发展动态。2. 多膜光纤声传感器采用普通的通信光纤和通用元器件制作开发的尝试。3. 对我国系列耳机发展提出了一个供与会代表研讨的初步构思和建议。

通信电声器件方面主要介绍了我国电话机用压电式换能器(包括送话器、受话器、振铃发声器的)引进情况及其技术水平;电子电话机用动圈式受话器的研制

情况和技术性能等。

这些论文的宣读和讨论,引起了与会代表们极大的兴趣和关注。大家认为这次会议开得紧凑、热烈,达到了预期的效果,这是与天津电声器材厂、内蒙古电声厂的精心安排、热情服务分不开的,代表们表示衷心感谢。

(王以真)