

第 五 届
全国语音图象通讯信号处理学术会议

论 文 集

THE 5TH NATIONAL CONFERENCE ON
SPEECH IMAGE COMMUNICATION
AND SIGNAL PROCESSING



中国电子学会 信号处理学会 语言图象通信专业委员会
中国声学学会 语言听觉和音乐声学分科学会

第五届全国语音图象通讯信号处理学术会议
(SICS'91)

1991年10月22日-25日

湖南 大庸

中国电子学会 信号处理学会 语音图象通信专业委员会
中国声学学会 语言听觉和音乐声学分科学会

前 言

中国电子学会信号处理学会语音图象通信专业委员会和中国声学会语言、听觉和音乐声学分科学会联合主办的第五届全国语音、图象、通讯和信号处理学术会议于 1991 年 10 月 22-25 日在湖南大庸召开。这是该学术会议首次在北京以外的地区举行。

本届学术会议由长沙铁道学院和北京邮电学院具体承办。长沙铁道学院负责会议的组织工作,北京邮电学院负责论文评审工作。会议征文得到国内同行的广泛响应和支持,共征收文摘 157 篇,经程序委会有关专家审阅,最后录用 119 篇作为会议正式论文发表,它们是:

语音识别	15	图象处理	33
语音合成	12	通讯技术	13
语音分析	9	神经网络及应用	9
语音编码	11	谱分析与 DSP 技术	7
听觉与音乐声学	10		

这些论文具有广泛的覆盖面,涉及了相应学科最具时代感的方面,反映了国内这些领域的研究现状和最新进展,许多论文具有创新的思想 and 观点,许多论文是长期踏实工作的成果,具有较高的参考和应用价值。相信通过这次会议交流,将有力促进我们的研究工作,扩大我们的学术影响。

这里,我们对论文作者、会议支持者和工作人员表示衷心感谢!

袁保宗 谷士文 诸维明

一九九一年十月

第五届全国语音图象通讯信号处理学术会议 (SICS'91)

1991年10月22日-25日 湖南 大庸

主办单位

中国电子学会 信号处理学会语音图象通信专业委员会
中国声学学会 语言、听觉和音乐声学分科学会

承办单位

长沙铁道学院
北京邮电学院

会议主席 副主席

袁保宗	北方交通大学
谷士文	长沙铁道学院
张家禄	中国科学院声学研究所

程序委员会主席 委员

诸维明	北京邮电学院
李昌立	中国科学院声学研究所
蔡德孚	中国科学院电子学研究所
齐士铃	中国科学院声学研究所
方至	中国科学院心理学研究所
黄泰翼	中国科学院自动化研究所
王延平	武汉大学
余崇智	南京大学
俞斯乐	天津大学
王仁华	中国科学技术大学
胡光锐	上海交通大学
欧阳景正	华南理工大学

秘书长

陈德池	长沙铁道学院
-----	--------

目 录

A: 语音识别

1. 树搜索算法用于孤立字语音识别 王淑春 吴善培 (1)
2. 汉语普通话声母的分类与识别 徐秉铮 邱 伟 (5)
3. 普通话孤立字全音节语音识别系统 童 健 杨行峻 (9)
4. 同音辨认汉语方言识别系统 钱志远 (13)
5. 普通话声母分类与感知网络辨认方法 欧贵文 (17)
6. 改善连续汉语语音识别性能的若干措施 林道发 杨家源 罗万伯 王跟东 (21)
7. 连续汉语人机对话系统的研究 王跟东 林道发 杨家源 (25)
8. 听觉的多元机理与语音识别 梁之安 郭元兴 (30)
9. 基于知识的汉语语句识别 郑元成 袁保宗 林碧琴 (34)
10. 多维马尔可夫语音识别数学模型 李 旭 陈尚勤 (38)
11. 多人通用孤立词模糊识别系统的研究 胡关锐 (42)
12. 应用人工智能方法提高汉语语音识别系统的识别率 徐秉铮 钟文青 (46)
13. 基于语音规则的汉语韵母识别方法 严普强 张 晶 (50)
14. 汉语单音节大词汇语音识别关键技术问题 李昌立 莫福源 陈 韬 (54)
15. 一个独立 CPU 的语音识别器 杜笑平 杨启纲 杨家源 (59)

B: 语音合成

1. 高自然度的语音合成广播系统 许 军 梁满贵 袁保宗 林碧琴 郑元成 (63)
2. 提高规则脉冲激励的语音合成质量 刘福文 倪维桢 王德隽 (66)
3. 普通话语句合成用的声调模型及声调协调规则 杨顺安 (70)
4. MPLPC 汉语实时合成系统 倪 宏 李昌立 莫福源 孙金城 (74)
5. 具有高清晰度的语音合成算法 韦 岗 欧阳景正 (78)
6. 一个实用的汉语语音合成系统 杨家源 贺德珏 苟大举 杨启纲 罗万伯 (82)
7. 合成言语自然度的研究 吕士楠 齐士铃 张家禄 (86)
8. 汉语有声校对系统 方 元 余崇智 张建明 (90)
9. 语音合成技术在声学语音学研究中的应用 吕士楠 齐士铃 周同春 (94)
10. KX-2 共振峰语音合成器 赵利君 赵志强 张双荣 刘志文 方 英 (98)
11. 广州话文语转换系统研制 龙庆华 金惠生 任 平 司徒锡康(103)
12. 关于粤语语音合成系统的建立 金惠生 龙庆华 任 平 司徒锡康(107)

C: 语音分析

1. 普通话的基频概率分布 胡兴慧 张家禄 (111)
2. 对文本作词的自动切分和语音合成 莫锦贤 周海涛 (115)
3. 基于短时谱最小均方误差估计的语音增强和剩余噪声衰减 曹志刚 郑文涛 (119)
4. 汉语普通话语音数据库 孙金城 陈希清 李昌立 莫福元 倪 宏 李 彤 (124)

5. 用微机和 DSP 加速板实现高分辨率数字化声图仪 何国建 孙金城 李昌立 (128)
6. 普通话音节中擦音响音的时长 周同春 陈 珊 (132)
7. 汉语普通话元音共振峰可变范围的研究 祖漪清 (136)
8. 汉语普通话音节时长变化的动态模式 曹剑芬 (140)
9. 一种基于自适应滤波的语音增强方法 沈亚强 曹志刚 (144)

D: 语音编码

1. Wavelet 变换与低比特率语音编码 刘家康 柯有安 林茂庸 匡镜明 (148)
2. 相关矢量量化器的设计与性能分析 彭 磊 徐秉铮 贺前华 (152)
3. 数字移动通信信道模型在语音编码中的应用
..... 匡镜明 P. Noll 符风雷 刘家康 (157)
4. 低速率语音传输中的信道—信道联合编码 杨劲松 许 军 张树京 袁保宗 (161)
5. 一种改进的码激励线性预测编码方法 符风雷 匡镜明 刘家康 (165)
6. 线性预测声码器激励的改善 崔慧娟 赵立东 诸庆麟 (169)
7. 采用动态比特分配的 16KB/S 子带编码系统研究(SBC-D) 刘振雷 赵尔源 (173)
8. 一种新的快速矢量量化搜索算法 陈 在 张方葛 (177)
9. SB-ADPCM 宽带编码系统的实现及其应用 沈永育 倪维桢 (181)
10. 突发性语音输入的 ATM 信元统计复用器的性能分析 沈雁蓉 丁 炜 (184)
11. 一种快速向量量化编码算法: 半搜索算法 黄为民 (188)

E: 图象处理

1. 基于模型的人脸智能图象编码的三维构造模型建立和运动参数估计的研讨 ...
..... 蔡德孚 (192)
2. 基于三角形分割的面部表情图像合成实验研究 时永霞 程 磊 蔡德孚 (196)
3. IIS 智能图象数据库管理系统的对象结构模型及其实现 莫庆远 袁保宗 (200)
4. 综合图形/图象信息处理系统环境的结构模型及其实现 莫庆远 袁保宗 (204)
5. 近 α -剪裁平均滤波和近双窗修正剪裁平均滤波 杨立瑞 丁润涛 (208)
6. 一个改进的 MeTZ 自适应滤波器 费耀平 谷士文 (213)
7. 一种用于图象处理的高性能多处理机系统 吴湘淇 刘庆良 李 峰 王瑞英 (216)
8. 自适应 DPCM 可视电话 孟 泉 张家谋 (220)
9. 384KB/S 会议电视视频编码计算机模拟研究 陈教芳 蔡德孚 (224)
10. V 氏图细化编码 王兆华 (228)
11. HDTV 信号的自适应亚取样压缩编码 张春田 范宏寅 (232)
12. 一种线填充算法 肖定如 (236)
13. 散斑干涉图象的自适应方向滤波及条纹恢复方法 杨亦兵 王厚枢 (239)
14. 数字图象的语义模糊性测度及增强处理 张仲颖 蔡学助 (243)
15. 噪声图象边缘增强的稳健 M 型滤波器 程存学 赵 雷 (247)
16. 自适应 M 滤波用于图象噪声平滑 蔡勤春 刘 钊 顾德仁 (251)
17. 医学显微图象中一种高效的分割技术 吴湘淇 陈 颖 唐尧明 (255)

18. 半色调灰度图象的压缩编码	戴善荣	(259)
19. 一种实时的红外图象分割算法	陈东彬 张淑琴	(263)
20. 一种序列图象二维帧间变换编码方法	吴镇杨 朱丹群	(267)
21. 可变阶曲面拟合深度图象表述算法的实现与完善		
.....	李则光 卢朝阳 徐洪青 陆心如	(271)
22. 投影条纹法的误差分析	李 凯	(276)
23. 基于 TMS320DSP 的装配视觉检测系统	国澄明 吴 涛 王兆华	(282)
24. 利用体模型描述三维曲面物体	张 超 袁保宗	(287)
25. 深度图象的表面分割新方法	唐四春 袁保宗	(291)
26. 手画建筑结构计算简图自动读入系统的特殊符号识别	陈练兵 谷士文	(295)
27. 一种自适应的规定直方图修正新方法	钟志刚 胡正仪	(298)
28. 地图图素识别技术研究	于 明 张总成 刘锡安	(302)
29. 用硬件实现对图象边缘的实时处理	张总成 王红岩	(306)
30. 真彩色图象处理在无损物理鉴定中的应用	徐谈林	(310)
31. 小波理论在图象处理中的一些应用	何 立 茹国宝 王延平	(313)
32. 细胞显微图象处理及参数的提取	张总成 韩 松 畅继武 张淑敏	(317)
33. 一种快速计算图象物体不变矩的算法	曾禹村 强 明	(321)

F: 通信技术

1. 一种基于 CSMA / CD 的优先级随机分割综合语音 / 数据 / 图象局部网协议 ...		
.....	莫庆远 袁保宗	(325)
2. 短波通信自适应选频系统与实时选频技术	李 旭 陈尚勤	(329)
3. 在数字电话通路上传输视频信号的新方法	周继成	(333)
4. 语音模拟置乱及其剩余可懂度	张知易	(337)
5. 电话线彩色静态图象数字通信系统	戴井刚 梁福成 刘厚祥	
.....	程 蟬 李晓明 李文敏	(341)
6. 改进电视制式中的信号处理	俞斯乐 何 璠	(345)
7. 短波高速数传定时新方法	段 涛 诸维明 李昌立	(351)
8. 格型自适应均衡算法的研究	郭 铭 吴德本 李昌立 莫福源	(355)
9. 基于 RSA 算法的保密通信协议设计与实现	李 岩 李怀诚	(359)
10. 实用数字式多频信号接收方法	梁满贵 袁保宗 王 泉 许 军	(363)
11. 2MB 终端机定时同步系统专用集成电路的设计	姚惠娟 朱月耕 冯重熙	(367)
12. 一种适合于 ATM 的 ADPCM 分组系统	郑宝玉 周芳培	(371)
13. 两 PUMA—562 机械手的协调运动	彭 军	(375)

G: 神经元网络及其应用

1. 一种新的人工神经网络模型	张立朋 蔡长年	(379)
2. 高阶 HOPFIELD 神经网络用于二维重叠物体形状识别	李 杰 谷士文	(383)
3. 一个非线性连续神经网络渐近稳定定理的推广	苗振江 袁保宗	(388)

4. 人工神经网络在多道时间序列模型自适应参数估计中的应用 成奇明 张树京 (392)
5. 基于神经网络的语音识别 陈永生 袁保宗 林碧琴 (396)
6. 神经网络用于自适应均衡和干扰抑制联合系统的研究 向增俊 毕光国 翁剑峰 (400)
7. 神经网络识别器鲁棒性研究 陈道文 (404)
8. 循环神经网络的语音识别算法 李海洲 徐秉铮 詹 剑 (407)
9. 用神经网络检测拓扑性质——探索视觉过程的多层次性 郭宝龙 郭 雷 (412)

H: 听觉与音乐声学

1. 智能化听力语言诊断与康复系统 杨道淳 方 元 余崇智 (416)
2. 听觉脑干电位的频谱分析 陈洪文 顾 瑞 于黎明 (420)
3. 计算机辅助颞骨内面神经全程结构重建及测量 戴 朴 姜泗长 顾 瑞 方耀云 王今著 (424)
4. 窄带噪声对大白兔听觉影响: 畸变产物声发射与脑干电反应测听对照观察 石勇兵 姜泗长 顾 瑞 (428)
5. 豚鼠不同强度脉冲声暴露后听反应阈及耳蜗病理变化实验观察 翟所强 王沛英 陈洪文 (432)
6. 短纯音诱发的 AP—SP 复合波和滤波短声诱发 AP 的比较 李兴启 孙建和 (435)
7. 听性脑干反应头皮分布图 张雁歌 于黎明 顾 瑞 (439)
8. 脉冲噪声作用后耳蜗电位早期变化的实验观察 郑杰夫 姜泗长 顾 瑞 李兴启 (443)
9. 听觉诱发电位的时变滤波叠加 于黎明 邵殿华 李兴启 顾 瑞 (447)
10. 喉全切除术后发音钮的研究及临床应用 刘清明 秦逢彬 孙 健 张桂珍 韩在文 张玉莉 王 坤 齐士铃 (451)

I: 谱分析与 DSP 技术

1. 通用实时信号谱分析系统 USSA-25E 杨鸿魁 陈锡先 尹建其 (455)
2. 用奇异值分解法对非高斯参数模型定阶方法的研究 陈红敏 黄绣坤 (459)
3. 非线性—线性自适应滤波 蔡勤春 史剑雄 顾德仁 (463)
4. 一种基于多结构元素的形状的形态分解方法 周晓琪 袁保宗 (466)
5. 一种用于语训仪的低成本快速信号处理器 杨子冬 陈道文 (470)
6. CORDIC 算法的 VLSI 实现及其在信号处理中的应用 钮心忻 孙允恭 (474)
7. 数字信号处理器 TMS320C30 开发系统的在线仿真软件 莫福源 李毅民 傅国健 (478)

树搜索算法用于孤立字语音识别

[Isolated-word Recognition Using Tree-search Algorithm]

王淑春

吴善培

Wang shu-chun

Wu shan-pei

(北京邮电学院 邮政编码 100088)

Beijing University of Posts & Telecommunication

[摘要]本文介绍了连续密度隐马尔可夫模型(HMM)多说话人孤立数字语音识别系统,以及我们提出的树搜索算法在孤立字语音识别中的应用。

一. 引言

语音识别的研究是人们用科学的方法去探索语音的产生和感知过程,是在机器理解人类语言的基础上,将语言以控制的形式表现出来。

语音识别目前有四个发展方向:从特定人到非特定人,从孤立字语音到连续语音,从小词汇量到大词汇量,从干净语音到含噪语音。认人的大词汇量的语音识别目前已有进入商用的。在大词汇量的孤立字语音识别中,提高识别速度格外重要,因此本文提出了一种树搜索识别算法,大大提高了识别速度。

1983年Rabiner等人将隐马尔可夫模型用于孤立字语音识别,在非特定人识别领域内取得了突破,是一种有一定吸引力的识别方法,在八十年代形成了一大热潮。因此本文也采用了这种方法。

二. 系统介绍

识别系统框图如图一所示:

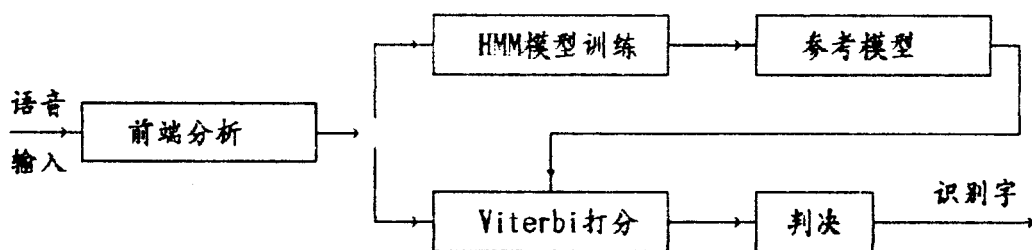


图 1. 识别系统框图

前端分析包括语音信号的采集、端点检测、分帧加窗予加重及特征提取,本系统采用如下规定:

采样率: 8KHZ

帧长: 20ms即160点

帧移: 10ms

窗函数: Hamming

予加重系数: 0.97

特征参数采用 10 阶倒谱系数和一阶能量归一化对数值。

模型选用连续高斯密度隐马尔可夫模型，马氏链选用左至右结构，状态数选为 5，初始模型选取如图 2 所示，训练时采用 Baum-welch 重估公式，如式 (1)、(2)、(3) 所示：

$$(1) \hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(x_{t+1}) \beta_{t+1}(j) / P_r(x/\lambda)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_{t+1}(i) / P_r(x/\lambda)}$$

$$(2) \hat{V}_j = \frac{\sum_{t=1}^T Y_t(j) \cdot X_t}{\sum_{t=1}^T Y_t(j)}$$

$$(3) \hat{U}_j = \frac{\sum_{t=1}^T Y_t(j) (X_t - V_j) (X_t - V_j)'}{\sum_{t=1}^T Y_t(j)}$$

其中 $Y_t(j) = \frac{\alpha_t(j) \beta_{t+1}(j)}{\sum_{j=1}^N \alpha_t(j) \beta_{t+1}(j)}$ $b_j(x) = N(x, V_j, U_j)$

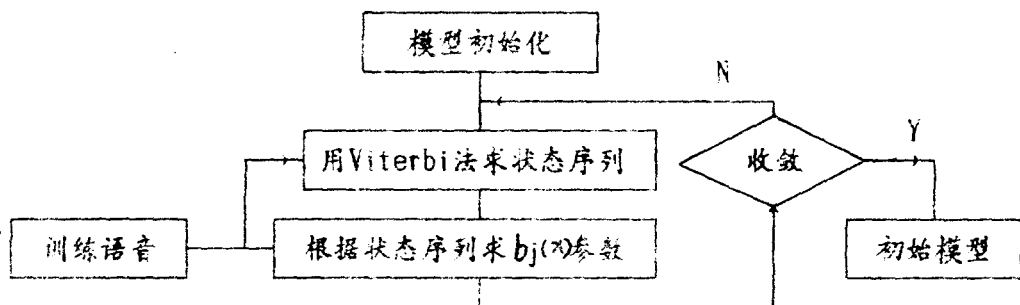


图 2 初值选取框图

三. 树搜索算法应用于语音识别

1. 树搜索概念

树搜索的概念首先由矢量量化中提出，它与全搜索相对应。在矢量量化中，所谓的全搜索矢量量化，就是在全矢量码本空间内，根据某一最近邻准则寻找某一矢量的最近邻矢量作为此矢量的量化值，而树搜索矢量量化，是在矢量码本空间中逐级搜索最终找到局部最近邻矢量作为此矢量的量化值。

树搜索有二叉树和二叉树之分，但它们的原理是相同的，下面以二叉树为例说明树搜索原理。

由二叉树的结构图(图3)可见，以树根(第 0 层)为起点，第一层有两个节点 (Y0, Y1)，第二层有四个节点 (Y00, Y01, Y10, Y11)，第 N 层有 2^N 个节点。

搜索步骤：首先由树根开始，分别搜索第一层中的各个节点，根据要求选择一节点，然后根据树结构由此节点决定搜索下一层节点的范围(如：第一层所选节点为 Y0，则第二层搜索范围为集合 {Y00, Y01})再在此范围内搜索找出所需节点，

依次类推直到树的最后一层。

很显然，树搜索能减少搜索次数，降低运算量。对于四层二叉树，最后一层的节点数为 16 个，树搜索只要搜索 8 次，而对于同样搜索空间 (16 个节点) 全搜索算法，却要搜索 16 次。

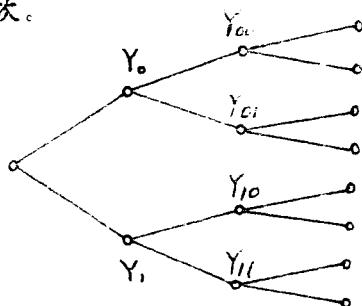


图 3 二叉树结构图

2. 树搜索算法应用于孤立字识别

汉语语音可以根据不同特征将语音分成不同类，如声调、发音和频谱等，因此可以利用树搜索法进行识别。

训练时，首先将语音根据某一特征进行分类，每类训练产生一个代表性的模型作为树的第一层的节点。然后再对每类中语音根据其特征再进行分类，然后再训练产生其代表性的模型作为树的第二层节点。依次类推，直到建立每个语音的模型。

识别时首先搜索第一层的节点，确定属于哪一节点，然后由树的结构确定第二层的搜索范围并进行搜索，确定属于第二层的那一节点，依次类推，直到搜索到最后一层找到其对应的识别字为止。

四. 实验结果

我们对孤立数字的语音识别进行了讨论。首先我们对十个孤立数字的识别情况作了详细的统计，如表一所示，这里的模型是 30 人每人 2 遍训练产生的，训练方法如前所述。

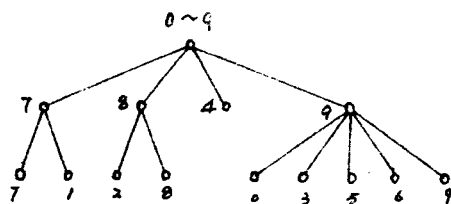


图 4 识别时的树结构

根据表 1 我们将十个数字语音分成四类，第一类含有 1 和 7，第二类含有 2 和 8，第三类含有 4，第四类含有 0, 3, 5, 6, 9。树结构如图 4 所示。由表可以看出：我们可以用 7, 8, 4, 9 的模型分别作为四个类的模型。这里我们所采用

的模型是 15 个人每人 2 遍建立的 20 个模型(男女各十个), 识别结果如表 2 所示。由表 2 可知, 平均识别率从 93% 变化到 89.1%, 但是由图 4 可以看出识别速度提高了 36%。

表 1 (表中的识别率都是百分制计)

	0	1	2	3	4	5	6	7	8	9
0	76.7	0	0	0	0	0	10	6.6	0	6.7
1	8.3	73.3	0	0	0	0	0	10.4	10.4	0
2	0	0	45	8.3	0	0	0	0	45	1.7
3	0	0	0	81.6	3.3	0	3.3	0	8.3	3.5
4	0	0	0	5.0	91.7	0	1.6	0	1.77	0
5	0	0	0	0	0	96.7	0	0	0	3.3
6	0	0	0	3.3	0	0	80	0	0	16.7
7	0	10	0	0	0	0	0	90	0	0
8	0	0	3.3	1.6	0	0	1.7	0	90	3.4
9	0	0	0	0	0	1.7	5	0	0	93.3

表 2:

	0	1	2	3	4	5	6	7	8	9
a	98.3	93.3	90	86.7	91.6	100	98.3	95	81.7	95
b	83.7	81	85	89	100	96.7	80	95	86.6	93.3

a 为全搜索识别结果

b 为树搜索识别结果

[参考文献]

1. 语音信息处理, 蔡长年, 北京邮电学院研究生教材。
2. Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities, L.R.Rabiner, B.H.Juang, S.E.Levinson, M.M.-sondhi, AT&T vol.64 No.6, 1985
3. Vector Quantization in Speech Coding, John Markoul, Salim Roucos, and Herbert Gish, Proc. IEEE vol.73 No.11 1985
4. High Performance Connected Digit Recognition Using Hidden Markov Models, L.R.Rabiner, Jay G.Wilpon and Frank K. Soong, IEEE ASSP vol.37 No.8 1989

汉语普通话声母的分类与识别

徐秉铮 邱伟

(华南理工大学 无线电与自动控制研究所)

Classification and Recognition of Chinese (Putonghua) Consonant

Bing-zheng Xu Wei Qiu

(Research Inst. of Elec. & Auto., South China Univ. of Tech.)

一. 汉语语音识别方案

在国内外汉语语音识别研究中,目前较成熟的是以全音节为单位来进行识别,这一方案充分利用了汉语语音以音节为发音单位这一特点,受协同发音的影响较少。但从识别的角度来看,数量仍然偏多,计算量较大,且音节之间的细微差别常被整个音节的相似性所淹没,不利于不认人的语音识别。

另一个国内研究得较多的是以音素为单位的语音识别,在汉语普通话里只有 32 个音素,如果能以音素作为识别单位,无疑有样本数和存储量极少的优点,不幸的是,音素之间的分割是一件十分困难的事,目前尚未找到一种有效的方法将它们准确地切分开来;另外,音素受协同发音的影响很大,由于以上二个致命的弱点,使音素为识别单位研究的进展不大,目前尚属探索阶段。

进一步研究汉语音节的特点,我们注意到,所有的汉语音节都可用一个公式表示,即:音节=声母+韵母+声调。在汉语普通话里,只有 22 个声母,38 个韵母及 5 个声调(包括零声调)。如果用声韵母作为一个音节的识别单位,所需样本数比全音节的要少好几倍,且它们受协同发音的影响比音素要小得多。另外,声、韵母的切分比音素要容易很多,同时,它对不认人连续语音识别也比较适用。因此,将声韵母分开识别作为一个由认人向不认人语音识别系统过渡的识别手段是较为理想的。

本文将讨论一种汉语声母实时识别的方案,这个方案已在本所研制的汉语声控打字机中得以实际应用。

二. 汉语声母识别策略

从频谱特性来看,除了浊声母 [m, n, l, r] 成谐波状,具有基频和共振峰外,其它声母动态频谱没有明显的共振峰特性、不存在周期性、频带宽、呈噪声特点且受后接韵母的影响大,因此给识别带来了不少困难。

有文献 [3] 报道利用声、韵母的过渡段信息来识别声母,对于有些声母,其区别特征并不完全由声母本身携带,而其后接的韵母也携带了一部分,这种现象称为“音征互载”,即区别信息存在于后接韵母相连的过渡段中,由于某一声母与不同韵母结合时,有不同的过渡段,所以一个声母要建立多个模板,据报道,考虑了“音征互载”效应后,声母识别率可达到 94%,但为此付出的代价是:声母模板数从原来的 21 个增加到 74 个,这种方法的另外一个问题是过渡段的划分;文献 [6] 巧妙地避开了这一难点,提出以声母加过渡段来识别声母的方法,也得到了较好的效果。

以上的方法不失为声母识别的好方法,但它们实现起来略嫌麻烦,且计算时间开支较大,本文提出一种声母识别策略,以求实时准确地识别声母。

既然从频域上分析声母有一定的困难,那么我们换一个角度,即从时域上来看一下它们彼此间有没有不同之处。有人曾对声母的音长作了测量 [3] [4],

发现,虽然一个汉字的读音长度差异较大,但声母的音长却比较稳定,且不太因人而异。这就提示我们有可能从时域上着手快捷又准确地将声母分成若干类,将识别范围大大缩小,然后再利用语音分布知识及频谱性质对类内几个声母进一步细分,这样就能减少许多声母因频谱上的差别不显著性而形成识别精度不高的困扰(因为一旦分类成功,则只要找出类内固定几个声母的识别方法即可)。这样,我们将声母识别分解为二个过程:先进行分类识别,然后再根据声调和韵母的识别情况对声母进行进一步的类内识别。

三. 声母分类的实现

在研究声母时,传统上有依发音部位和依发音方式二种分类,若把数据依发音部位来归纳统计,数据离散很大,很难看出彼此的关系[4]。若把数据依几种发音方式归纳统计,数据离散很小,彼此构成一定比例关系。有人曾利用发音方式将声母分为六类,据报道分类准确率达92%。[5]

经过对声母发音方式的分析及对声母实验数据的仔细分类,并结合本系统实时性和高效性且向不认人方向发展的要求,我们提出了一种声母分类的有效方案:在众多的声母声学参量中抽取声母音长;平均过零率;声母信号的周期性;前五帧的平均过零率及相对能量等五个特征参量作为分类特征,将声母分为如下八大类:

<1> [零声母]; <2> [b, d, g] 不送气塞音; <3> [l, m, n, r] 浊音; <4> [z, zh, j] 不送气塞擦音 <5> [p, t, k] 送气塞音; <6> [f, h] 擦音; <7> [c, ch, q] 送气塞擦音; <8> [s, sh, x] 擦音

由于五个特征参量全部可从时域信号中直接得到,无需作任何变换,所以实现方便,计算简单且快速,可实时进行分类。

(1) 声母音长分析

文献[4]对声母的音长作了研究,得出22个声母的平均音长及其标准偏差的分布数据,音长按不送气塞音[b, d, g],不送气塞擦音[z, zh, j],浊音[l, m, n, r],送气塞音[p, t, k],送气塞擦音[c, ch, q],擦音[f, h, s, sh, x]这一顺序逐渐加大,男女声基本相同。我们在实验室请了四男三女发音,也得出类似的结果。

在代表音长的时间轴上,每类音各占有一定的区域,有些类间区域相互有重叠区,而有些类则互不重叠,如不送气塞音[b, d, g]和[零声母]的音长不超过30ms,擦音音长最短也有90ms,它们一定不会重叠。所以利用声母音长这一特征可以把一些类与类分开,也可粗略地确定出待识别声母所在类的范围。

(2) 声母平均过零率分析

短时平均过零率为我们提供了一个快速的谱估计分析,擦音、塞擦音的平均过零率最高(其中[f, h]又较低些),送气塞音次之,浊音最低(不送气塞音由于用音长参量即可分出,无需再求它们的平均过零率),利用平均过零率可以将擦音、塞擦音和其它声母分开。

(3) 声母信号的周期性分析

声母的发音根据声带振动或不振动分为浊声母和清声母。浊声母包括[零声母]和[l, m, n, r],其它的声母都是清声母。浊声母是由周期性声门波激励产生声带振动而成,其时域波形呈显著的周期性变化,清声母的时域波形中看不到周期性的变化,利用浊声母这一周期特性很容易将它们同清音分开。

(4) 前五帧平均过零率分析

在研究平均过零率时,我们注意到平均过零率的标准偏差比较大,这说明

帧与帧之间的平均过零率变化较大,进一步的研究发现,这种变化主要出现在声母的头几帧及声韵母过渡段上,送气塞音 [p, t, k] 和擦音 [f, h] 的前五帧平均过零率较其它擦音及塞擦音小,这个参量的主要作用是将送气塞音 [p, t, k]、擦音 [f, h] 和其它擦音及塞擦音区分开来。

(5) 前五帧相对能量分析

声母的发音过程可分为:成阻、持阻和除阻,在声母的除阻与接到韵母之间存在着一段过渡音,根据成阻的结构,持阻的持续和除阻的延伸以及过渡音音征变化这四段情况不同便形成各种发音方法,即产生各个不同的声母语音信号。由此可见,前几帧声母的能量变化即反应了成阻到持阻这一过程。

擦音 [s, sh, x] 和 [f, h] 成阻时发音部位两点相接近,但并不太紧,持阻时气流由发音部位的两点间所留下的窄缝间挤出,发生摩擦成声,所以相对能量较低。

塞擦音 [z, zh, j] [c, ch, q] 的成阻到持阻的前段呈塞音状,即成阻时发音部位的两点紧闭,持阻时保持这种阻碍,但到持阻的后段阻碍的部分放松了一些,使气流透出,变成擦音的摩擦,其发音是先强后弱,反映在前五帧的相对能量比擦音大。

[p, t, k] 为送气塞音,在时域上成阻和除阻的时间很短,但发音时透出一股较强的气流,反映在前五帧的相对能量最大。

根据前五帧相对能量的大小,就可以将送气塞音和擦音、塞擦音分离开来。

$$\text{前五帧相对能量 } E_n = \frac{E_5}{E_n + 2\sigma_{E_n}}$$

其中: E_5 : 前五帧能量总和;
 E_n : 背景噪声平均能量;
 σ_{E_n} : 背景噪声能量方差;

(6) 声母分类器的实现

根据前述的五个特征,我们着手设计一个高效的声母分类器(图略),将声母分为八大类。通过对十几个人的统计分析得出不认人分类的准确率达90%;认人分类的准确率可达98%。

四. 利用语音知识实现声母的二次识别

每种语言都有它自己独特的结构,汉语也是如此。在本系统中,我们考虑建立一些简单的语音知识库以达到进一步识别的目的。在汉语普通话里共有22个声母,38个韵母及四个声调(不计零声调),理论上它们可以组成3344个音节,但实际上汉语有音节字只有1282个(带零声调),这就提示我们,其中的2·3组合是不存在的,例如,声母分类得到 [p, t, k], 韵母识别结果为 /ong/, 声调识别结果为第二声,根据声、韵、调的组合情况可知不存在第二声 pong 和 kong 的音,则我们即可确定声母只能是 t, 在这种情况下,只利用语音知识就能唯一地从分类集合中挑选出声母;又如,分类结果为 [l, m, n, r], 韵母识别结果为 /ai/, 声调为第三声,则根据语音分布知识可知不存在第三声 lai 和 rai 的音,则可将声调分类的范围缩小为 [m, n], 需对它们作进一步的识别,但识别的范围仅通过语音分布知识就缩小到了二个,不但可减少进一步细识别的运算时间,而且提高了细识别的准确率。

我们从《现代汉语词典》中统计得出语音知识分布表(略),以声母和韵母作为检索路径,用一个字节中的5个比特分别表示0~4声调的分布情况。

/1/ 表示该声调存在, 0/ 表示不存在。例如在 /gai/ 栏中可查到 /01011/, 表示第一、三、四声存在, 而零声调和第二声不存在。

五. 用模式匹配法实现声母的细识别

利用语音分布知识固然可能将声母分类的范围进一步缩小, 有时甚至可唯一地从分类集合中挑选出声母识别结果, 但并不是对所有的语音都有同样的效果, 这主要取决于分布知识的本身, 在不少情况下, 声母的分类集合的范围并不能进一步缩小, 或范围虽缩小但并不能唯一地确定, 例如上节我们提到的第二例, 利用语音分布知识我们已将声母的范围从 $[m, n, l, r]$ 缩小到 $[m, n]$, 然而若想进一步识别究竟是 m 还是 n 就要采用其它有效的方法, 在本系统中, 我们利用模式匹配法来达到最后确定声母的目的。

在本系统中, 由于我们在此之前已用时域的几个特征参量对声母作了分类, 因此, 对声母的细识别的关键是要找出类内集合中各声母的细微差别, 我们考虑一种简单的方法, 以便充分利用我们的分类结果。根据我们对声母发音特点的分析发现: 声母信息的差别主要集中在二个部分, 第一部分为声母发音从成阻到持阻段, 即声母的前段; 第二部分差别则出现在除阻到后接元音之间的过渡段。我们正是利用了第一部分的差别来实现类内声母细识别的, 具体的做法是: 从声母起始点开始取五帧声母语音信号求15阶倒谱系数, 帧移长度根据声母的长度作动态调整, 五帧信号作时间归正后作模式匹配, 由于是类内几个声母的判别, 所以用这种方法一般都能有效地从类内集合中对声母作出进一步的识别。但对一些发音特点相似且发音者主观上都无法判别该发哪一个的声母, 如 /c, ch/, /s, sh/ 和 /z, zh/, 是无法进一步细识别的, 除非应用语音分布知识能将它们挑选出来, 否则, 系统将把这一模糊性交后续的词库检索来解决。

六. 结论

在声母的识别上, 本系统采用二级识别法, 使运算速度和匹配精度达到最佳。先用五个时域特征将声母分类, 然后用语音分布知识进一步缩小分类中元素的范围 (有时仅到这一步就可唯一地识别出声母), 最后, 用模板匹配法最终确定识别结果。本文所阐述的方法已在实际系统中投入应用, 声母识别率为95%以上, 是一种简便高效的声母实时识别的方法。

参 考 文 献

- [1] L.R.Rabiner and R.W.Schafer, Digital Processing of Speech Signal. Englewood Cliffs, NJ: Prentice-Hall, 1978
- [2] Xu Bingzheng, Du Minghui and Qiu Wei, "Chinese Speech Controlled Typewriter", ICSP'90, pp.505-508, Oct. 1990.
- [3] 陈永彬, 王仁华, 语言信号处理, 中国科技大学出版社, 1990
- [4] 齐士铃, 张家禄, 汉语普通话辅音音长分析, 声学学报, Vol.7, No.1, 1982年第一期.
- [5] 吴立忠, 汉语普通话辅音分类的研究, 华南理工大学硕士论文, 1986.
- [6] 高惠琴, "用1/3音节作匹配基元研究汉语单字音识别", 电声技术, pp.1-5, 1989 第一期.
- [7] 倪立民, 语言学概论, 浙江大学出版社, 1988.

普通话孤立字全音节语音识别系统

A Speaker Dependent Recognition System of the PuTongHua

童 健 杨 行 峻

Jian Tong XingJun Yang

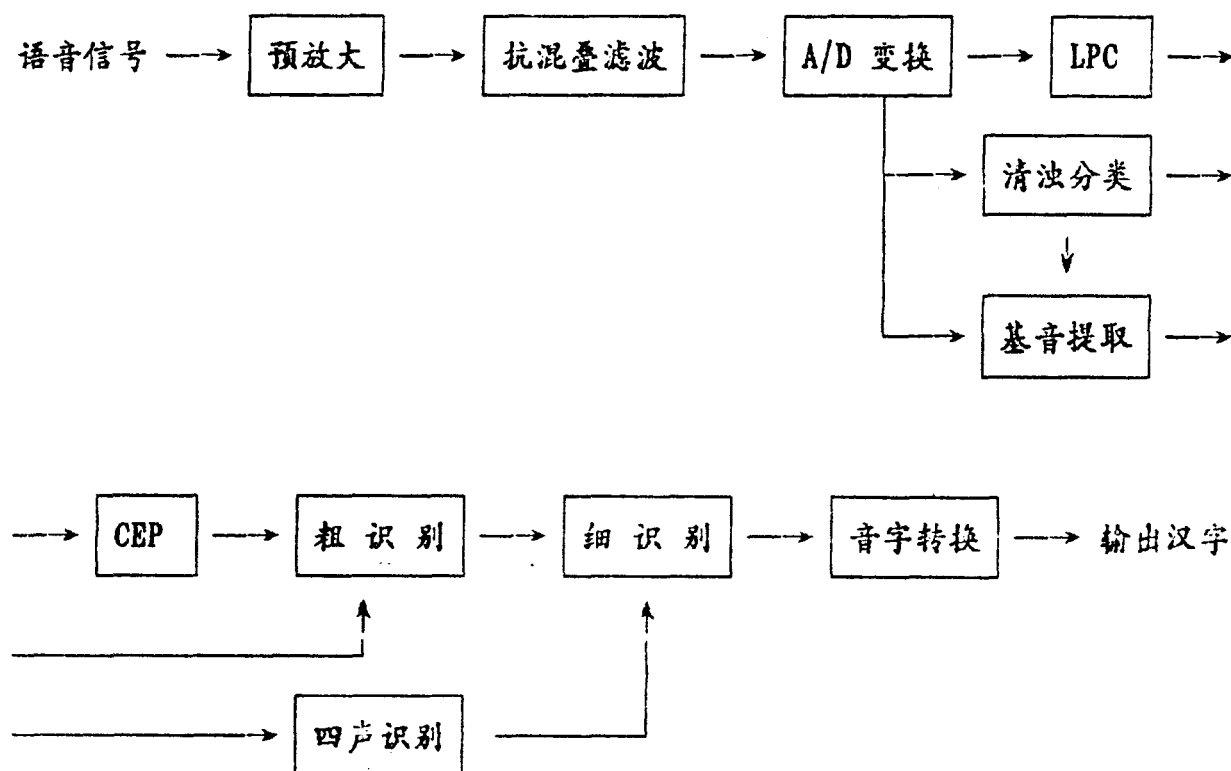
(清华大学电子工程系)

(Dept. of Electrical Engineering, Tsinghua University)

一. 概述

本文介绍一个汉语全音节语音识别系统, 该系统能实时识别汉语398个拼音音节及四声, 通过与汉字操作系统的接口, 实现汉字的计算机语音输入, 该系统由清华大学电子工程系开发, 并已由清华大学科技开发总公司组织生产销售。

系统构成如图一所示, 输入语音经放大, 抗混叠滤波后, 送至A/D变换器量化, 再经过语音参数提取(LPC系数, CEP系数, 基音周期, 能量等), 然后进行四声识别以及基于清辅音音长的声母分类和基于浊音匹配的韵母粗识别, 选出不超过40个候选音节, 最后进行音节细识别, 输出拼音及四声, 并将其转换为汉字。



图一 THASR汉语语音输入系统