

第 六 届
全国语音图象通讯信号处理学术会议

论 文 集

THE 6TH NATIONAL CONFERENCE ON
SPEECH IMAGE COMMUNICATION
AND SIGNAL PROCESSING 1993



NANPING · SICHUAN

1993 · 9

中国声学学会 语言听觉和音乐声学分会
中国电子学会 信号处理学会 语音图象通信专业委员会
四川省电子学会

第六届全国语音图象通讯信号处理学术会议 (SICS' 93)

1993年9月24日—26日

四川 南坪



赠 93.12.

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

前 言

中国声学学会语言、听觉和音乐声学分会和中国电子学会信号处理学会语音图象与通信专业委员会联合主办的“第六届全国语音图象通讯信号处理学术会议”于一九九三年九月二十四日至二十六日在四川南坪召开。这是语言声学、图象处理、信号处理及听觉领域中最受热烈响应的学术会议之一。

本届会议由电子工业部西南通信研究所(四川成都)承办,承担全部会务工作,由中国科学院声学研究所负责论文的征集、评审、出版等事宜。他们为办好这次会议付出辛勤的劳动,对开好会议做出了巨大的贡献,我们代表会议向他们表示衷心的感谢。

这次论文征集受到了国内同行的热烈响应,共征集论文126篇。经过程序委员会的专家评审,选取了91篇做为会议发表论文。它们包括以下几个方面:

语音识别	18篇	语音增强	3篇	听 觉	13篇
语音合成	9篇	图象处理	19篇	心理声学	1篇
语音分析	8篇	通信技术	8篇	生物声学	1篇
语音编码	3篇	信号处理	8篇		

这些论文具有广泛的覆盖面,既反映了国内主要研究单位的学术成就,也体现了国际间该领域的主要研究方向。许多论文是具有创新意义的,不少是当前信息革命浪潮中科技成果的结晶。它们既有学术价值又有应用意义。没有入选会议的论文也是有很好的价值、有先进的成果,仅仅是因为会议能容纳的人数及时间受限。欢迎下次再来参加会议。在这里我们也向所有向会议投稿,并做出努力的作者们表示感谢。

语音、图象、通信领域中信号处理是最有用武之地的领域,特别是当前时代进入信息革命,通信走向全球化,声象图文等多媒体信息处理发挥了从事通信、计算机、自动化和微电子专业的全部聪明才智的历史阶段,它更是具有领头性质的研究学科。这次会议交流的论文希望能在这样的方向上起到一些指路的作用。

最后,让我们为开好这次大会共同做出努力。谢谢参与会议的全体人员。

袁保宗 齐士铃 诸维明

一九九三年七月二十七

第六届全国语音图象通讯信号处理学术会议 (S I C S ' 9 3)

1993年9月24日—26日 四川 南坪

主办单位

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

承办单位

西南通信研究所

会议主席

袁保宗 北方交通大学

副主席

刘村友 西南通信研究所

程序委员会委员

张家骥	中国科学院声学研究所
诸维明	北京邮电学院
程 蝉	西南通信研究所
方 至	中国科学院心理学研究所
顾 瑞	解放军总医院
徐秉铨	华南理工大学
蔡德孚	中国科学院电子学研究所
侯朝焕	中国科学院声学研究所
黄泰翼	中国科学院自动化研究所
方棣棠	清华大学
苏曼波	西南通信研究所
齐士铃	中国科学院声学研究所

A: 语音识别

第六届全国语音图象通讯信号处理学术会议论文集目录

大会报告

- | | |
|-------------|-----|
| 1. 通信保护动向趋势 | 刘村友 |
| 2. 多媒体通信技术 | 程 蝉 |

分组报告

A: 语音识别

- | | | |
|--------------------------------|-----------------|----|
| 1. 汉语文本读入中音字转换的知识集成模型和时间同步搜索算法 | 徐近需 高 枚 | 1 |
| 2. 共振峰在语音分析识别中的研究 | 吴自强 马建华 宁 军 | 5 |
| 3. 用于非特定人全音节语音识别的神经网络方法研究 | 李广杰 方棣棠 李建民 | 9 |
| 4. 汉语声母识别研究 | 陈 韬 莫福源 李昌立 | 13 |
| 5. 不送气塞擦音声母J和送气擦音声母X识别 | 欧贵文 | 17 |
| 6. LSP参数与CEP参数在语音识别中的性能比较 | 陈 翔 余崇智 方 元 | 21 |
| 7. 用矢量量化方法进行说话人确认的参量研究 | 邱小军 吴阿华 | 25 |
| 8. 语音识别应用多段VQ时两级量化算法研究 | 刘承望 兰国梁 鹿凯宁 李爱军 | 29 |
| 9. 汉语单音节中声母识别系统的研究 | 王 敏 宁 军 吴自强 马建华 | 33 |
| 10. 神经网络汉语数字串识别 | 李海峰 徐近需 | 37 |
| 11. 非特定人汉语全音节语音识别的研究 | 李建民 方棣棠 | 41 |
| 12. 汉语全音节声母分类器的研究 | 宁 军 吴自强 王 敏 马建华 | 45 |
| 13. 汉语语音分析识别实验系统 | 吴自强 马建华 王 敏 宁 军 | 49 |
| 14. 基于模糊统计的非特定人数字音语音识别 | 黄亦青 余崇智 方 元 | 53 |
| 15. 过零周期转移概率矩阵语音识别方法的改进 | 张连海 韩 华 王炳锡 | 57 |
| 16. 战斗机座舱环境下语音识别系统的实现 | 杨子云 徐近需 王承发 | 61 |
| 17. 采用80C552单片机的声控拨号电话 | 杨子云 | 65 |
| 18. 汉语声韵自组织映射分类图及在语音识别中的应用 | 陈道文 黄泰翼 马 茜 马宗强 | 69 |

B: 语音合成

- | | | |
|---------------------------|-----------------|-----|
| 1. 汉语语音合成协同发音规则初探 | 唐涤飞 吕士楠 周同春 王仁华 | 75 |
| 2. 基于CELP的汉语语音规则合成 | 倪 明 华一满 | 79 |
| 3. 汉语大词汇合成系统的研究 | 倪 宏 李昌立 莫福源 孙金城 | 83 |
| 4. 汉语语音合成中的多音字处理方法 | 谭志强 徐近需 | 87 |
| 5. 汉语韵母音节的共振峰参数的提取与合成算法研究 | 周俏峰 蔡莲红 | 91 |
| 6. 藏语声调感知在合成中的应用 | 孔江平 | 95 |
| 7. 汉语规则合成的字音流自动转换 | 李 彤 莫福源 陈 韬 | 99 |
| 8. 藏文语音合成管理系统 | 孔江平 徐 昂 | 103 |
| 9. 汉语轻声音节合成规则研究 | 初 敏 吕士楠 周同春 | 107 |

C: 语音分析

1. 论语音声学参数数据库	鲍怀超	111
2. 汉语的数学模型	陈芳 袁保宗 林碧琴	115
3. 汉语词汇语音统计特性初探	孙金城 倪宏 李昌立 莫福源	119
4. 儿化韵习得中介实验研究	朱川	123
5. 普通话塞音音节中元音共振峰模式及其协同发音的研究	颜景助	127
6. 汉语普通话双音节中V ₁ -b, d, g声学表现的实验研究	陈肖霞	131
7. 普通话客观音质评价的研究	吴淑珍 吴阿华	135
8. 改进型LPC残差法提取语音基音周期	程启明	139

D: 语音编码

1. 线谱对(LSP)及其在低速语音编码中的应用	夏光荣	143
2. 语言信号的信源/信道联合编码	王耀武 陆以勤 韦岗	147
3. 高质量音乐信号实时编码系统	韦岗 熊鹰 欧阳景正	151

E: 语音增强

1. 听觉模型与语音增强	陆生礼 余崇智	155
2. 用主分量分析法对含噪语音的分析	方元 杨道淳	159
3. 一种利用倒谱消除混响的方法	李国峰 张松	163

F: 图象处理

1. 64Kb/s 活动可视电话硬件系统设计	王敬 李坚 诸维明 张家谋	167
2. 大容量文档图象的MMR压缩的软/硬件实现技术	蔡德孚 刘建	171
3. 纹理边缘检测与随机分形理论在图象压缩中的应用	苟飞 余英林	175
4. 层叠滤波图象边缘检测研究	丁润涛 高峰	179
5. 深度图象的快速分割	邓世伟 袁保宗	183
6. 视频图象压缩算法及实现研究	陈敦芳 翟宝光	187
7. 算术码在图象数据压缩编码中的应用	戴善荣	191
8. 二维DP在图象模式识别中的应用	叶军 刘大奔 李海洲	195
9. 一种减少双目立体视觉匹配特征的算法—近视原理	陈俊华	199
10. 分类分层遥感图象数据库	唐四春 阮秋琦	203
11. 一种利用局部形状信息实现细化畸变校正的方法	张树声	207
12. 基于Mallat算法的图象压缩方法	房育栋 余英林	211
13. 多证据合成的指纹识别方法	黄为民 边肇祺	215
14. 一种用于方向分解编码的二维低频内插方法	范宏寅 张春田	219
15. 运动信息处理的神经推理系统	郭宝龙 郭雷	223
16. 一种利用视觉特性和小波变换的混合图象编码方案	冯为民 阮秋琦 袁保宗	227
17. 基于模型的结构变形法人脸图象编码技术	梁慧颖 蔡德孚	231
18. TMS320C50图象处理系统	国澄名 刘军 王兆华	235
19. 面向对象的方法在图象信息系统中的应用	王传军 俞志和 朱志刚 徐光佑	239

G: 通信技术

1. 两种2.4kb/s 声码器噪声环境适应性的研究	张知易 王春生	245
2. 移频编码在水声图象通信中的应用	蒋 晖 侯朝焕	249
3. 高性能FSK纠错编码信号解调系统—MCS-51 单片机在水声通信中的应用	李平安 王秀杰	253
4. ISDN终端 LAPD一致性测试系统的理论与实现	唐蔚云 杜 森	257
5. 图象通信中的差错控制技术	戴井岗	261
6. 一种可变阶的数字相关器	曾兴雯 杜武林 尹伟谊	265
7. ISDN U-接口中的新型线路均衡网络	蒋 华 李 正	269
8. 伪随机噪声法在回声对消、信道均衡上的应用	马骏驰 郭明阳	273

H: 信号处理

1. 客运列车智能式语音自动报站报时系统	陈德池	277
2. 中文文本压缩的自适应算法	贺前华 徐秉铮	281
3. 基4FFT专用IC结构设计	胡国荣 祝裕斌 侯朝焕	285
4. 超级对象系统的研究	张 军 袁保宗 尹作友	289
5. 多媒体应用软件开发平台	王跟东 杜笑平 费向东 朱庆华	293
6. 自适应窗堆栈滤波器及脉冲噪声滤波	程存学 刘长贵	297
7. 通用语音信号处理系统的实现	程启明 张礼和	301
8. 利用RBP网络重构吸引子	范 统 徐秉铮	305

I: 听 觉

1. 长时噪声暴露对频率选择性的影响	方 至 王乃怡	309
2. 振动、噪声与脉冲声复合因素对豚鼠耳损伤的实验观察	程所强 陈洪文 王沛英 顾 瑞 姜泗长	313
3. 耳蜗弱音电位(CM)起源的探讨—豚鼠急性缺氧实验研究	李兴启 孙建和 孙 伟 姜泗长	317
4. EOAE输入—输出关系与耳蜗非线性	于黎明 顾 瑞	321
5. 豚鼠耳蜗电图慢成分研究	王树峰 刘淑芳 姬柏春 邓元诚	325
6. 语言信号处理中的耳蜗宏观力学模型	徐世钰 戴善荣	329
7. 畸变产物耳声发射的基本特性及与瞬态诱发性耳声发射的关系	宋本波 刘 博 刘 铤	333
8. 自发性耳声发射与畸变产物和瞬态诱发性耳声发射的关系	宋本波 刘 博 刘 铤	337
9. 感音神经性聋的几种耳声发射监测	刘 博 宋本波 刘 铤	341
10. 耳鸣与耳声发射关系初探	刘 博 宋本波 刘 铤	345
11. 耳声发射测试系统传感装置的研究	翁立忠 林金森 张正国 张建一 王建明	349
12. 聋哑儿童听力补偿的研究	林金森 张伍玲 张正国 周 亦	353
13. 新一代可视语言训练系统(VSTS)	杨子冬 陆 嚶 陈道文	357

J: 心理声学

1. 词边界的声学语音学线索 杨玉芳 361

K: 生物声学

1. 蝉类用于声通讯的鸣声特性及其飞行趋声范围的估计 蒋锦昌 杨新宇 唐欢 徐慕玲 陈浩 365

基于分维度 CELP 分形码本的构造 陆以勤 王耀武 韦岗 369

汉语文本读入中音字转换的知识集成模型和时间同步搜索算法

徐近濡 高校
哈尔滨工业大学

一. 概述

计算机语音输入的目的是把输入的语音序列转化为正确的文字序列。解决此问题的关键在于综合运用各种可用的声学知识和语言学知识和约束来消除音字转换中的歧义。

本文提出一种集成声学知识和语言学知识的组合概率模型, 并将此模型用于音字转换过程的动态规划搜索算法, 可以时间同步方式综合运用全部知识于路径生成和选择, 找到最佳的前N个候选词串。文中还初步验证了所提出的模型和搜索算法的有效性。

二. 语音识别和音字转换的组合概率模型。

如有一声学发现矢量序列 $\underline{A} = a_1 a_2 \dots a_T$, 可用某种分割得 $\underline{A} = A^{(1)} A^{(2)} \dots A^{(N)}$, 其中每一声学子段 $A^{(n)}$ 可产生多个候选词 $W_1^{(n)}, W_2^{(n)}, \dots, W_{N_i}^{(n)}$, 其声学发现概率分别为 $P(A^{(n)} | W_i^{(n)})$, $n=1, \dots, N_i$ 。

给定声学发现 $\underline{A}$ 后最可能的词序列 $\underline{W} = w_1 w_2 \dots w_T$ 可由下得到:

$$\underline{W} = \underset{\underline{W}}{\operatorname{argmax}} P(\underline{W} | \underline{A})$$

最大后验概率

$$= \underset{\underline{W}}{\operatorname{argmax}} P(\underline{A} | \underline{W}) \cdot P(\underline{W})$$

Bayes规则

$$= \underset{\underline{W}}{\operatorname{argmax}} P(\underline{A} | \underline{S}) \cdot P(\underline{S} | \underline{W}) \cdot P(\underline{W})$$

两级解码 (图1-b)

$$= \underset{\underline{W}}{\operatorname{argmax}} P(\underline{A} | \underline{S}(\underline{W})) \cdot P(\underline{W})$$

$$P(\underline{S} | \underline{W}) = \begin{cases} 1 & \text{if } \underline{S} = \underline{S}(\underline{W}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

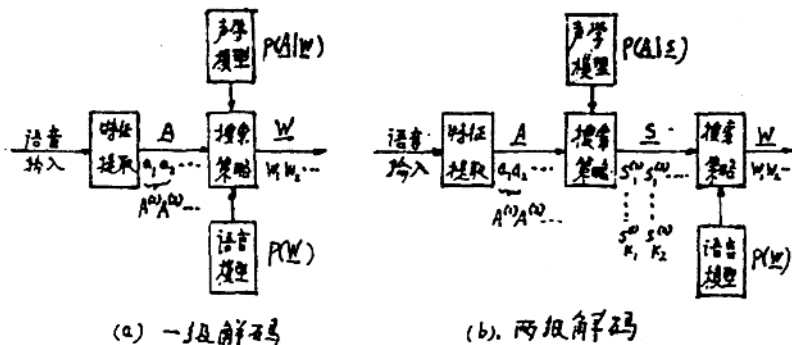


图1 语音识别系统框图。

其中:

1. $P(\underline{A} | \underline{S}(\underline{W}))$ —— 声学发现概率

是在采用两级解码 $\underline{A} \rightarrow \underline{S} \rightarrow \underline{W}$ 时, 先由声学级利用声学模型 $P(\underline{A} | \underline{S})$ 将声学发现 $\underline{A}$ 分割和识别为语音单元 (音节或词拼音) 的拼音串 $\underline{S}$, 再由语言模型 $P(\underline{W})$ 将其转化为词串 $\underline{W}$ 。当每一

词 w_i 有唯一对应的拼音串 $S(w_i)$ 时,则 $P(A|S(W))$ 即为产生词串 W 所对应的拼音串 $S(W)$ 时,发现声学序列 A 的概率。假设各词发音相互独立,则

$$P(A|S(W)) = \prod_{i=1}^I P(A^{(i)}|S(w_i))$$

其中 $P(A^{(i)}|S(w_i))$ 可由声学级提供的在声学子段 $A^{(i)}$ 产生拼音候选 $S(w_i)$ 的声学评价 $KS(w_i)$ 求得:如在 i 时刻的候选为 $(S(i, i, k), KS(i, i, k)), k=1, K$,则对单音节点

$$P(A^{(i)}|S(i, i, k)) = KS(i, i, k) / \sum_{k=1}^K KS(i, i, k) \quad (2)$$

对多音节点

$$P(A^{(j+1)}|S(j, i, k)) = [\prod_{l=j}^i P(A^{(l)}|S(l, l, k_l))]^\alpha \quad 1/(j-i+1) \leq \alpha \leq 1 \quad (3)$$

2. $P(W)$ ——词串 W 出现概率

应综合反映句中各词间的近邻远邻约束。采用组合概率模型

$$P(W) = P_b(W) \cdot P_w(W) \quad (4)$$

其中 $P_b(W)$ ——双词性(bi-POS)模型,反映近邻词性提供的近邻(局部)约束

$$P_b(W) = P(w_1|U_1^{(1)}) \prod_{i=2}^I P(w_i|U_i^{(1)}) \cdot P(U_i^{(1)}|U_{i-1}^{(1)}) \quad (5)$$

$P(w_i|U_i^{(1)})$ —— w_i 在词性 $U_i^{(1)}$ 中的发生概率

$P(U_i^{(1)}|U_{i-1}^{(1)})$ ——相邻词性 $U_{i-1}^{(1)}, U_i^{(1)}$ 的关联概率

$P_w(W)$ ——概率文法(Probabilistic Grammar)模型,反映句中所有词间的远邻(全局)约束。

利用上下文无关文法的CNF形式:

$$(g) U^{(1)} \xrightarrow{P(g)} W \quad (6)$$

$$(r) U^{(2)} \xrightarrow{P(r)} BC \quad (7)$$

$$P_w(W) = \prod_{j=1}^J P(r_j) \quad \text{if } S \xrightarrow{\{r_1, \dots, r_J\}} W \quad (8)$$

其中(g)型规则概率因已包含在 $P_b(W)$ 中,此处不再考虑。

则得音一字转换的组合概率模型:

$$\bar{W} = \arg \max_{\bar{W}} [P(A^{(1)}|S(w_1)) \cdot P(w_1|U_1^{(1)}) \cdot \prod_{i=2}^I P(A^{(i)}|S(w_i)) \cdot P(w_i|U_i^{(1)}) \cdot P(U_i^{(1)}|U_{i-1}^{(1)}) \cdot \prod_{j=1}^J P(r_j)] \quad (9)$$

采用负对数得相加的代价(距离)量度:

$$\begin{aligned} \bar{W} = \arg \min_{\bar{W}} D(A, \bar{W}) &= \arg \min_{\bar{W}} [-\ln P(A|S(\bar{W})) \cdot P(\bar{W})] \\ &= \arg \min_{\bar{W}} [D(S(w_1)) + D(w_1|U_1^{(1)}) + \sum_{i=2}^I [D(S(w_i)) + D(w_i|U_i^{(1)}) \\ &\quad + D(U_i^{(1)}|U_{i-1}^{(1)})] + \sum_{j=1}^J D(r_j)] \end{aligned} \quad (10)$$

其中有四类距离:

(1). 声学距离

$$\text{单音节 } D(S(i, i, k)) = -\ln P(A^{(i)}|S(i, i, k)) \quad (11)$$

$$\text{多音节 } D(S(j, i, k)) = 1/\alpha \sum_{l=j}^i D(S(l, l, k_l)) \quad (12)$$

(2). 词的类内距离

$$D(w_i|U_i^{(1)}) = -\ln P(w_i|U_i^{(1)}) \quad (13)$$

(3). 词性间联结距离

$$D(U_i^{(1)}|U_{i-1}^{(1)}) = -\ln P(U_i^{(1)}|U_{i-1}^{(1)}) \quad (14)$$

(4). 句法规则距离

$$D(r_j) = D(BC|U^{(2)}) = -\ln P(BC|U^{(2)}) \quad (15)$$

三. 知识组织和搜索算法

上述组合概率模型提供了将声学评价、词频、词性关联、句法等多种约束综合用于评价候选词串的数学模型。需要构造一种快速搜索算法,以上述模型为指导,从读入的语音基元候选阵列所产生的海量候选词串中尽快找到信度最高并合乎语法的前 N 个候选句子。计算前提为:声学级以单音节方式读入,产生候选音节阵列,各列可不等长,但应包含正确候选:可由单音节构成1-5音节的多字词拼音;各拼音可能有多个同音字;同一词可能有不同词性。因此,算法将面对海量候选词串。

3-1. 知识组织

系统将语言学有关知识组织于下列几个知识库中。

1. 词表: 将词按词长分别存于1-5字词表中, 各词项数据结构为

($g, w, S(w), U^{(1)}, D(w | U^{(1)})$)

2. 规则库: 用于存储(r)型规则, 各规则项数据结构为

($r, U^{(2)}, B, C, D(BC | U^{(2)})$)

3. 词性联结距离库: 用于存储前后词性间联结距离矩阵, 其数据结构为

($U_{1-1}^{(1)}, U_{1-1}^{(2)}, D(U_{1-1}^{(1)} | U_{1-1}^{(2)})$)

3-2. 时间同步的动态规划搜索算法

搜索算法的基本思想是: 对输入的候选音节阵列(图2-a), 以音节为时间单元, 由左到右, $i=1, (+1), I$, 用时间同步方式, 对每个时刻 i , 进行下列1、2操作。然后作回溯操作。

1. 词生成: 根据候选音节的可能组合, 利用词表信息生成1-5音节组成的 (j, i) 时段内所有可能词, 计算其代价(包含声学距和词类内距), 并为每个时段 (j, i) 、每个词性 $U^{(1)}$ 保留代价最小的前 K 个候选的词号和打分。

2. 路径生成: 利用规则库找到一直到句子始端的所有 (j, i) 时段中合乎规则(7)的所有剖析子树(图2-b), 计算其代价(包含左、右子树的代价, 左、右子树相邻词的词性联结距和规则距), 并为每个时段 (j, i) 、每个生成属性 $U^{(2)}$ 保留代价最小的前 K 个候选路径和打分。

3. 回溯: 当搜索到句子终止 $i=I$ 时, 如所生成路径中有生成属性为句子起始符 S , 且代价小于规定门限 SD_0 的路径时, 则可用回溯方法找到合乎句法的前 N 个最佳候选句子。

因篇幅所限, 略去具体算法。

四. 实验和讨论

为检验所提出的模型和算法的可行性和有效性, 用小型语料库进行了实验。

4-1. 实验条件

采用有关科学和政治的两篇文章作为建立知识库的依据。在词表中补充了与语料中的词同音和近音的较常用词及常用的功能词。分别建立了1-5字词表, 共含词条1800余项。词性按语法为主、语义为辅划分为26种, 再附加短语、句分属性12种。按此, 建立了含有91条组合规则的规则库, 其中规则概率简单地采用二值决策, 即对所有合乎语法规则的组合均取其发生概率为1, 即规则距为0。

用上述语料统计产生了词的类内概率 $P(w | U^{(1)})$ 和词性联结概率 $P(U_{1-1}^{(1)} | U_{1-1}^{(2)})$, 并按此算得类内距 $D(w | U^{(1)})$ 和联结距 $D(U_{1-1}^{(1)} | U_{1-1}^{(2)})$, 存入词表和联结距离库。由于统计语料太少, 因而参照 H.Ney 提出的方法[4] 以均匀分布作非线性内插对统计结果进行概率平滑。

实验语句选用语料中13个句子, 其平均句长为12个字。输入拼音数据取自我们研制的“汉语文本读入系统”的输出, 它以候选音节阵列的形式存入文件, 每项以(拼音、打分)表达, 每列平均候选数为8个音节, 并必含正确音节在内。图3-a例示一个以单字方式读入的句子“化学是研究分子变化的”的候选音节阵列。在我们所用词表和所选例句中, 每个输入单字最多可有16个同音或近音字。为解决单音节同音近音字过多问题, 实验中采用 $\alpha = 1/(j-i+1)$, 以保证声学评价为长词优先。

整个实验用C语言在IBM-PC/AT上实现。

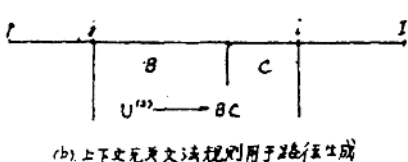
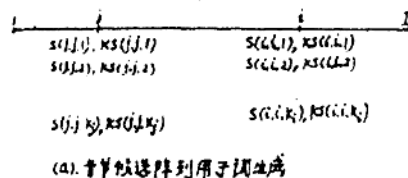


图2. 搜索算法中词和路径生成

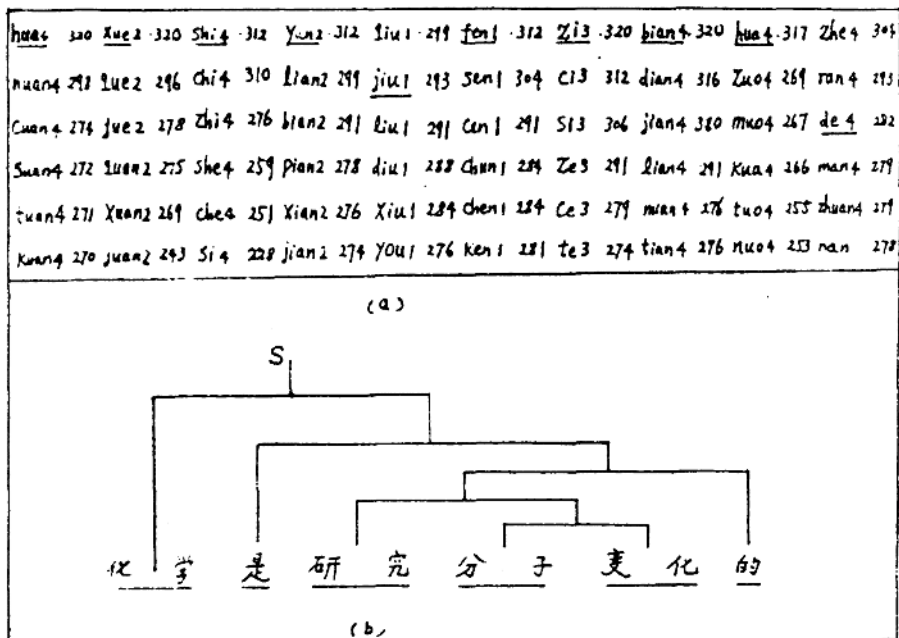


图3 音字变换示例

4-2. 实验方案和结果

为检验模型中各类知识的有效性,在上述实验条件下比较了下列三种方案:

- (1). 用全部知识作搜索。
- (2). 用无近邻约束的其余知识作搜索。
- (3). 用无近邻和远邻约束的其余知识作搜索

用上述三种方案对13个例句的候选音节阵列作音字变换搜索,并输出前5个候选句子及代价。可以得出:(1)中有12个正确句子在第一候选,有一个在第三候选。图3-b示例由图3-a候选音节阵列得到的句子剖析树。(2)中则有11个正确句子在第一候选,有2个不在前五个候选中。(3)中则只有7个正确句子在第一候选,而其余6个均不在前五个候选中。由此可见综合近邻和远邻约束知识对于正确路径的搜索和排序的重要性。

4-3. 讨论

上述结果初步揭示了所提模型和算法的可行性,以及综合各类语言学知识对于正确实现音字转换的重要性。进一步工作是:

1. 在大语料库基础上考核模型和算法的有效性。
2. 加速计算以满足实时性要求。
3. 研究容错的词汇搜索和句法剖析算法,以适应输入未含正确音节及文法未覆盖情况。

参考文献 (略)

Knowledge-Integrated Model and Time-Synchronously Search Algorithm for Speech-Character Transcription in Chinese Text Read-in System

Xu Jinpei Gao Mei
 Harbin Institute of Technology, China

共振峰在语音分析识别中的研究

Formant Research in Speech Analysis and Recognition

吴自强 马建华 宁军
Wu Ziqiang Ma Jianhua Ning jun

国防科技大学电子技术系407教研室
Lab. 407, Dept. of Electronics Engineering, National University of Defense Technology

【摘要】 在本文中,作者根据实时识别系统的要求,从语音LPC分析入手,研究了语音共振峰提取的基本原理。分析了提取共振峰的主要难点,用理论与实验相结合的方法,找到了用BH多项式求根法所需要的最佳LPC阶数,提出了一种新的提取共振峰的修正算法。并给出了作者提取共振峰的部分实验结果(打印曲线)。利用共振峰参数已经比较好地解决了韵母识别及全音节语音合成问题。在语音匹配识别中,根据语音共振峰特性,提出了一种新的匹配距离定义。

【关键词】 语音识别,共振峰,语音LPC分析

语音共振峰是描述语音最本质的参数(特别是浊音),共振峰代表了发音信息最直接的来源^[1],具体应用在语音的分类鉴别和韵母识别以及共振峰语音合成。

求共振峰的方法很多,主要归结为以下几类:语图仪法^[11],峰值检测法^[2],多项式求根法^[3],卡尔曼滤波法等。与其他几种方法比,多项式求根法预测的共振峰参数精度高,分辨率高,用该算法与伪峰剔除准则相结合,可以确保提取全部共振峰参数。考虑到我们系统的实时分析并对参数有较高的精度要求,我们采用的是多项式求根法提取共振峰特征参数。

一:求根法提取共振峰的基本原理:

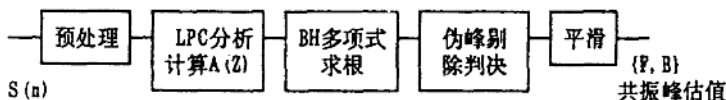


图1 多项式求根法实现框图

求根法的基本思想是:先对语音进行LPC分析,得到声道传递函数 $(1/A(Z))$,然后采用劈因子(BH)多项式求根法,求多项式 $A(Z)$ 的根,可以推出共振极点 Z_k 和 Z_k^* 与共振峰 (F_k, B_k) 存在以下关系^[10, 11]:

声道的共振峰带宽:

$$B_k = - \left[\frac{\ln |Z_k|}{\pi T} \right]$$

声道的共振峰频率:

$$F_k = \frac{1}{2\pi T} \arctg \left[\frac{\text{Im}(Z_k)}{\text{Re}(Z_k)} \right]$$

二:求根法提取共振峰的几个要点

1. 确定LPC分析参数

(1). 采样率和分析帧长:

由于部分消音的高频成份可达6~8KHZ,因此,系统的采样率设计为16KHZ。在保证声道具有平稳的统计特性(元音:15~20ms),系统分析帧长选16ms,帧覆盖为8ms。

(2). 预加重

一般预加重的方法^[9]是将信号通过只具有一个零点的滤波器 $1-\mu Z^{-1}$, ($\mu=0.9\sim 1$),该预加重模型也可等效为一个非线性高通滤波器。由于上述等效高通过于简单,实际应用中无法完全抵消声门波和口唇幅射的作用,同时对低频(接近零频)信号衰减过大,这对于提

取鼻音和介音类的母第一共振峰非常不利。因此,作者在系统设计时,去掉了预加重网络,实验效果令人满意。如图2、图3所示。

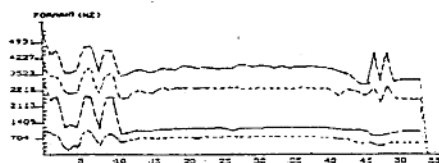


图2: 预加重对“Mang”音共振峰的影响

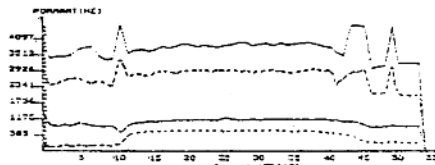


图3: 去掉预加重“Mang”音共振峰曲线

(3). LPC预测阶数

选择模型LPC分析阶数的一般原则: 首先保证足够多的极点来模拟声道响应的谐振结构。根据发声机理的分析^[6], 以及共振峰语音合成^[12]的实验结果, 语音声道完全可用5个共振峰来模拟其特性^[8]; 5个共振峰对应5对极点, 即需10阶LPC分析。在上述模型中, 由于没有考虑声门和唇部的幅射特性, 故 $P=10$ 是最低限。考虑声门波频谱有 $-10 \sim -18\text{dB/倍频程}$ 的滚降, 而唇部幅射特性一般近似为 $+6\text{dB/倍频程}$ 的提升^[10], 1971年J. D. Markel提出一种提取共振峰所需LPC阶数的计算方法^[9]: f_s (采样率KHz) 加上4或5。这个数值是根据实验结果而得出, 对于声道平均长度为 17cm 和声速 34cm/ms 来说, 相当于附加4或5个极点。但是, 作者通过大量各种阶数共振峰参数提取实验发现, 对我们目前的语音分析系统($f_s=16\text{kHz}$, 无预加重网络)来说, f_s 加4至5个点的LPC阶数, 并不完全合适, 应加以修正。作者认为实际的LPC分析阶数应为10阶加4到6阶较为合适。10阶为提取声道五个共振峰起码必须具备的, 4至6阶是为抵消实际声道中出现零点或声门波激励和幅射的组合效应; 实际取值是16。后经大量实验表明, 用16阶LPC提取韵母的共振峰效果是令人满意的。

2. 多项式求根算法

提取共振峰参数, 首先必须求解声道传递函数的极点, 也就是说要解一个一元16阶方程, 并求其所有的根(实根、复根)。一般方法有: 牛顿-拉夫逊(Newton-Raphson)搜索算法; 半区间搜索法; 劈因子BH(Bairstow-Hitchcock)算法。本系统采用BH算法^[9, 4], 求解 $A(z)$ 多项式的根。

3. 求根法提取共振峰参数的修正判决

由于本系统LPC分析阶数较高, 使得这些共振峰原始数据中存在伪峰; 如何“去伪存真”, 这一直是一个比较难的问题。作者利用语音学的先验知识, 分析和提出了一套有效的逻辑判决准则。对提取的共振峰进行修正、筛选, 实验的结果表明, 效果是令人满意。

基本思想和算法:

(1). 根据语音学知识, 我们知道, 语音共振峰前五峰的范围一般为 $250 \sim 6000\text{Hz}$ ^[7]。由于我们系统未加预加重网络, 声门波中的低频成分对共振峰产生影响, 在共振峰原始数据中, 有时存在 $0 \sim 100\text{Hz}$ 的峰点。作者设计了一个判决门限, 当 $f < 100\text{Hz}$ 时, 该数据作为野值剔除。

(2). 语音共振峰的带宽的定义为: 语音声道每一个二阶谐振器谐振频率对应的幅值下降3dB的频带宽度。语音学的知识告诉我们, 共振峰前五峰的带宽一般为十几Hz~1kHz左右^[7]。因此, 对于共振峰的修正可以利用带宽信息加以判断。当共振峰带宽 $>1500\text{Hz}$ 时, 作者认为是无意义的峰值数据, 予以剔除。

(3). 当相邻共振峰带宽之比大于6, 而频率之差(绝对值)小于 150Hz 的数据, 判为伪峰, 予以剔除。

(4). 对修正后的共振峰数据进行排序处理。

4. 共振峰数据的平滑

本系统采用中值平滑法,对最后修正的数据进行平滑。这种方法采用一个沿数据滑动的窗口,用窗口样点的中值取代窗口中点的数值,这样可消除共振峰数据中较小的不规则性,而不会使明显的非连续性变得模糊不清。实验结果是比较满意的,例如作者发“ai”音:

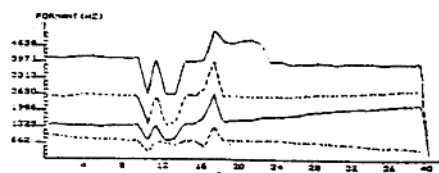


图4: “ai”修正前共振峰曲线

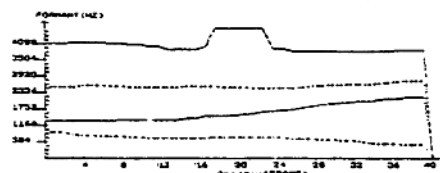


图5: “ai”修正后共振峰曲线

三. 共振峰提取的实验结果

采用上述共振峰提取算法,作者作了大量语音分析实验(近3000个音);另外,采用该算法,还进行了共振峰语音合成实验^[9],其结果都是令人满意的(尤其是对于韵母)。如下图,值得一提的是,采用该算法我们建立的汉语全音节共振峰参数库。

附部分共振峰实验曲线图:

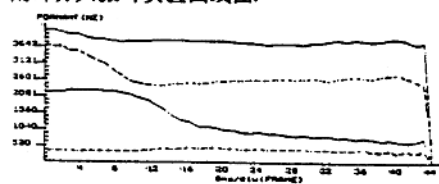


图6: 复合韵母“iou”

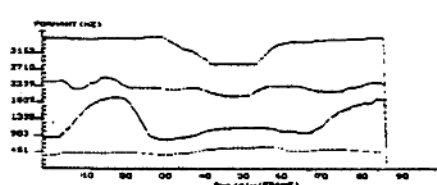


图7: 作者发音“we were away”
(可与文献^[2]比较)

四. 共振峰语音匹配识别

利用共振峰很好地解决了韵母的识别问题;同时对含非擦类声母的全音节,识别效果也较好。系统采用改进DTW匹配方法,分别用绝对值距离、带状距离,对35个韵母进行匹配实验,实验结果如下:

(1) 绝对值距离: 权系数: $w(i) = \{5, 1, 1\}$, 识别正确率: $35/35=100\%$

(2) 带状距离:

由于共振峰特征在小范围有一定的波动性,我们提出一种带状距离,假设用共振峰前3峰进行匹配:

设 F_T 为待识别语音的特征矢量; F_R 为库中语音的特征矢量;

带状距离的定义为:

$$d_c(F_T, F_R) = \sum_{i=1}^3 d(F_T(i), F_R(i))$$

$$d(F_T(i), F_R(i)) = \begin{cases} e_i & |F_T(i) - F_R(i)| > \Delta f_i \\ 0 & \text{else} \end{cases}$$

实验选择: $\Delta f_i = \{100, 100, 200\}$ $i=1, 2, 3;$

$e_i = \{2, 1, 1\}$ $i=1, 2, 3;$

上述距离定义的主要优点是允许待识别语音特征在一定范围(Δf_i)的波动;同时,可使匹配的区别度增大。

带状距离匹配实验结果:

权系数: $w(i) = \{1, 1, 1\}$; 其余条件同上:

首选识别正确率: 34/35=97.1% [ing] 错匹配为[u].

前二候选识别正确率: 35/35=100%

结果分析:

①. 只要共振峰参数求的稳定, 匹配可以得到相当大的区别度, 如e的第一、二候选距离之比接近十万倍. 这对于提高识别、分类的性能是非常好的方法.

②. 匹配区别度可通过调整 $\Delta f_i, w(i)$ 人为控制.

③. 计算速度快.

这也许是用共振峰特征进行匹配的最佳距离.

【参考文献】

- [1] T. W. Parsons, "Voice and Speech Processing", McGraw-Hill Book Company 1986.
- [2] N. B. Pinto, D. G. Childers, "Formant Speech Synthesis: Improving Production Quality", IEEE Trans. ASSP, Vol. 37, No. 12, pp1870-1884.
- [3] J. D. Markel, A. H. Gray, "Linear Prediction of Speech", Jr. Springer - Verlag Berlin Heidelberg 1976.
- [4] 《常用算法》科学出版社, 1976
- [5] L. Lamel, L. R. Rabiner, A. E. Rosenberg and J. G. Wilpon, "An Improved Endpoint Detector for Isolated Word Recognition", IEEE Trans. ASSP, Vol. Assp-29, No. 4, Aug. 1981
- [6] 王成友 "汉语语音合成系统" 国防科大电子技术系硕士论文, 1993. 1
- [7] 吴宗济、朴茂灿主编 《实验语音学概要》高教出版社, 1987
- [8] J. D. Markel, "Formant Trajectory Estimation from a Linear - squares Inverse Filter Formulation", SCRL Monograph No. 7, Speech Comm. Research Lab., Santa Barbara, California (1976).
- [9] 《计算技术通讯》1975年1-2期.
- [10] L. R. Rabiner, R. W. Schafer, "Digital Processing of Speech Signals." Prentice-Hall Inc., 1978.
- [11] 吴宗济 《汉语普通话单音节语图册》中国科学院出版社