

Research on Speech recognition and
Speech Visualization

语音识别及 语音可视化技术研究

■ 韩志艳 著

 东北大学出版社
Northeastern University Press

语音识别及语音可视化技术研究

Research on Speech recognition and Speech Visualization

韩志艳 著

东北大学出版社

· 沈 阳 ·

© 韩志艳 2017

图书在版编目 (CIP) 数据

语音识别及语音可视化技术研究 / 韩志艳著. — 沈阳: 东北大学出版社, 2017. 1
ISBN 978-7-5517-1539-3

I. ①语… II. ①韩… III. ①语音识别—研究 IV.
①TN912.34

中国版本图书馆 CIP 数据核字(2017)第 032657 号

内容简介

本书系统地研究了语音识别与语音可视化问题。全书共分为 9 章。第 1 章介绍了课题的国内外研究现状、意义和需要解决的难题,第 2 章对语音信号的预处理技术进行了介绍,第 3 章介绍了语音静态特征参数提取方法,第 4 章介绍了语音动态特征参数提取方法,第 5~7 章介绍了基于小波包变换和 K-L 展开的语音识别方法、基于小波混沌神经网络的语音识别方法、基于遗传算法和 Fisher 投影的语音识别方法等的研究成果,第 8 章介绍了语音可视化技术,第 9 章归纳全书并对今后工作提出展望。

本书的主要特点是在语音识别和可视化等方面提出了开创性的设计和分析方法,书中的内容来源于作者近年来的创新性研究成果,新颖实用,研究方法先进,尤其注重语音识别和可视化算法的鲁棒性和实用性。

出版者: 东北大学出版社

地址: 沈阳市和平区文化路三号巷 11 号

邮编: 110819

电话: 024-83683655(总编室) 83687331(营销部)

传真: 024-83687332(总编室) 83680180(营销部)

网址: <http://www.neupress.com>

E-mail: neuph@neupress.com

印刷者: 沈阳航空发动机研究所印刷厂

发行者: 东北大学出版社

幅面尺寸: 170mm×240mm

印 张: 14

字 数: 304 千字

出版时间: 2017 年 1 月第 1 版

印刷时间: 2017 年 1 月第 1 次印刷

责任编辑: 孙 锋 李 佳

封面设计: 潘正一

责任校对: 图 图

责任出版: 唐敏志

ISBN 978-7-5517-1539-3

定 价: 50.00 元

前 言

语音是语言的声学表现，是人类交流信息最自然、最有效、最方便的手段，也是人类思维的一种依托。而对聋哑人来说，语言交流变成一件很难实现的事情。一部分聋哑人不能说话是因为他们的听觉器官遭到破坏，不能将语音信息采集到大脑，但发音器官是完好的。这种情况下的聋哑人，如果辅助于一些视觉训练系统，经过一段时间的专门训练，是可以学会说话并和健全人进行交流的，这样为聋哑人进行听力无损补偿的语音可视化技术便应运而生。本书便立足于这一研究构想，通过提取语音信号的特征参数，将其与图像进行映射，产生具有声音意义的图像，供聋哑人学习并认知，辅助聋哑人听到声音。而语音信号特征提取是关系到语音识别和可视化系统性能的一个重要指标，目前提取的语音特征参数在安静的环境下具有很好的鲁棒性，但是这些参数一旦应用于噪声环境时，其性能会急剧下降。所以本书主要针对低信噪比环境下特征参数的提取及这些特征参数在语音识别和语音可视化中的应用进行了深入的研究。

本书的主要研究内容和创新点有以下几个方面。

(1) 为了提高低信噪比下语音端点检测的准确率，提出了一种新型的语音端点检测算法。该算法利用短时能量法和鉴别信息的互补优势，一边降噪一边端点检测，而且对于被判为噪声的帧利用基于子带能量鉴别信息方法来进行二次复检，根据鉴别信息来更新噪声能量门限，从而能更准确地跟踪噪声的能量变化。实验结果表明，该方法在信噪比变化比较剧烈的情况下仍然能够准确地进行端点检测，对语音信号的后续处理起到了很好的作用。

(2) 动态特性是语音多样性的一部分，它不同于平稳的随机过

程，它具有时间相关性，揭示了语音信号前后及相邻之间存在着的密切关联。由于差分参数和加速度参数并不能将动态信息挖掘得很充分，所以，它们尚不能很好地反映语音信号的动态特性。而调制谱具有时频集聚性，它不仅充分地反映语音之间的动态特性，而且对语音环境的敏感度较低。所以，本书根据干扰信号与语音信号在调制信息中不同的反应，提取调制信息中有效的语音成分，然后与 MFCC 参数的提取方法类似来提取其倒谱特征。这样得到的特征参数鲁棒性更好。

(3) 提出了一种基于共振峰曲线的语音信号动态特征提取方法，采用基于 Hibert-Huang 变换的方法来估算预处理后的语音信号共振峰频率特征，然后按照从第一帧到最后一帧的帧序，将预处理后的每帧语音信号的第一共振峰频率特征值进行组合获得第一共振峰曲线，依此类推，获得第二共振峰曲线、第三共振峰曲线及第四共振峰曲线。对获得的每条共振峰曲线进行快速傅里叶变换获得线性频谱，然后再求取能量谱，计算对数能量和离散余弦变换。与现有方法相比，本书提取的语音信号动态特征具有时间相关性，揭示了语音信号前后以及相邻之间存在着的密切关联，提高了语音识别的性能。

(4) 本书利用小波包变换和 K-L 展开对 0—9 十个数字语音进行了识别，该方法首先对语音信号进行预处理，其中运用倒谱距离检测算法对语音信号进行了端点检测，然后通过小波包变换提取出 32 个子频带的相对能量与语音信号的总能量一起作为原始的特征向量，然后通过 K-L 展开将 33 维的原始特征向量降为 6 维，通过建立 BP 网络分别用 33 维特征向量和 6 维特征向量对数字语音信号进行识别，实验结果显示该方法在降低特征参数维数的同时还保持了较高的识别率，有很好的识别效果。

(5) 基于语音信号的时变特性，本书提出了一种新型神经网络语音识别方法——小波混沌神经网络方法，即把小波变换和混沌特性引入到神经元，构成小波混沌神经网络，把这种神经网络用于语音识别，并与常用的 BP 神经网络识别方法进行了比较。实验结果表明，小波混沌神经网络的平均识别率要高于同等条件下常用的神经网络方法的识别率。

(6) 提出了一种基于遗传算法 (GA) 和 Fisher 投影的最佳可鉴

别基的求解方法。将原始特征向量向着最佳可鉴别基投影可得到具有最佳可分性的新的特征向量。本书以 6 个汉语元音信号为实验数据，以该信号的 MFCC 倒谱系数、小波变换的子带能量比、基音周期、共振峰频率和短时过零率作为原始的特征向量，使用遗传算法找出了原始特征的次优组合，并组成新的特征向量，将 GA 找出的次优特征向量向着 fisher 最佳可鉴别基投影可得到最佳鉴别特征向量。最后用混沌神经网络 (CNN) 作为分类器得到了较好的识别效果。

(7) 基于聋哑人的视觉鉴别能力和对色彩刺激的视觉记忆能力较强的优点，提出了四种可视化方法：第一种是基于局部线性嵌入 (LLE) 和模糊核聚类相结合的方法，先采用本书提出的改进的 LLE 对特征进行非线性降维，然后再利用模糊核聚类算法对其进行聚类分析，即利用 Mercer 核，将原始空间通过非线性映射到高维特征空间，在高维特征空间中对语音信号特征进行模糊核聚类分析。由于经过了核函数的映射，使原来没有显现的特征凸显出来，从而能够更好地支持基于位置的语音可视化，经过实验验证具有很好的效果。第二种是基于集成特征和神经网络的可视化方法，通过集成不同的语音特征进入一幅图像中为聋哑人创造了语音信号的可读模式。首先对语音信号进行一系列预处理，然后提取其特征，其中用经过正交实验设计优选的 23 个特征送入神经网络 II 映射出位置信息，用三个共振峰特征来对图像的主颜色信息进行编码，用声调特征来对图案信息进行编码，最后合成出可视化图像。我们对该可视化系统进行了初步的测试，并与以前的语谱图方法进行了比较，测试结果表明：该方法应用在聋哑人辅助学习方面可以收到良好的效果，具有很好的鲁棒性。第三种是提出了一种基于颜色和图案的汉语声韵母可视化方法，首先通过对语音信号进行一系列预处理，然后提取其特征，其中用共振峰特征来映射图像的主颜色信息，音长来映射图像宽度信息，谐振强度来映射图像长度信息，再将所有特征参数全部送入神经网络训练，神经网络的输出将会映射出图案信息。以汉语的声母和韵母为研究基元，对该可视化系统进行了初步的测试，测试结果表明：该方法具有很好的鲁棒性，对单独声母和韵母的正确回答率为 94.56%，两拼音为 85.75%，三拼音为 78.05%。第四种

是提出了一种基于神经网络的汉语声韵母可视化方法，采用小波神经网络来进行位置信息映射和颜色信息获取，由于小波神经网络具有结构可设计性、收敛精度可控性和收敛速度快的优点，有效地提高了汉语声韵母的正确编码率。而且将图像分为 12 个不同颜色的显示区域，每个显示区域内的音具有相似的发音特点和相同的发音部位，这就更好地利用了聋哑人对色彩刺激的视觉记忆能力较强的优点。与现有方法相比，具有很好的鲁棒性和易懂性。

本书讲述的内容为作者近年来的研究成果，内容新颖，属于当前所属研究领域的前沿问题，具有重要的理论与应用价值。

在本书的编写过程中，渤海大学工学院尹作友教授、王健博士等提出了许多有价值的建议，硕士研究生张鹏、夏桂阳、徐静、殷允飞在手稿整理、仿真实验和校对书稿等方面做了大量的工作，作者在此向他们表示由衷的感谢。

本书的出版获得了国家自然科学基金（项目编号：61503038、61403042）和辽宁省教育厅一般项目（项目编号：L2013423）的资助，在此一并表示感谢。

由于作者水平有限，不足之处在所难免，热诚欢迎读者与同行不吝赐教。

韩志艳

2017 年 1 月于渤海大学

目 录

第1章 绪 论	1
1.1 语音信号研究背景概述	1
1.2 国内外研究现状	3
1.2.1 语音识别技术研究	3
1.2.2 语音可视化技术研究	4
1.2.3 语音信号特征参数提取技术研究	6
1.3 课题的研究意义	7
1.4 课题研究需要解决的难题	8
1.5 章节安排	9
第2章 语音信号预处理技术	13
2.1 概 述	13
2.2 语音生成系统和语音感知系统	13
2.2.1 语音生成系统	13
2.2.2 语音感知系统	15
2.3 语音信号生成的产生模型	18
2.4 语音信号的时域波形	20
2.5 语音信号的采样和量化	22
2.6 语音信号的预加重	23
2.7 语音信号的分帧和加窗处理	23
2.8 语音信号端点检测技术	25
2.8.1 短时能零积法	26
2.8.2 信息熵法	27
2.8.3 频带方差法	28
2.8.4 HMM 模型法	29
2.8.5 倒谱距离测量法	30

2.8.6	基于 DWT 的互相关函数法	31
2.9	一种低信噪比下的语音端点检测算法	33
2.9.1	降噪方法	33
2.9.2	子带能量的计算	34
2.9.3	鉴别信息的计算	34
2.9.4	算法描述	35
2.9.5	实验结果对比及分析	36
2.10	本章小结	41
第 3 章	语音静态特征参数提取	46
3.1	问题的提出	46
3.2	短时能量	47
3.3	短时平均幅度	47
3.4	短时过零率	47
3.5	短时自相关函数	48
3.6	短时平均幅度差函数	48
3.7	倒谱特征	49
3.7.1	基于线性预测的倒谱参数 LPCC	49
3.7.2	基于 Mel 频率的倒谱参数 MFCC	50
3.8	感知线性预测参数	52
3.9	共振峰特征参数	53
3.10	谐振强度特征参数	54
3.11	基于单边自相关序列的语音特征	56
3.12	基于独立分量分析的鲁棒语音特征	57
3.12.1	模型描述	58
3.12.2	频域独立分量分析	58
3.12.3	特征提取算法描述	61
3.13	基于线谱频率参数的语音特征提取	69
3.13.1	线谱频率参数分析	70
3.13.2	求解线谱频率参数的常用算法	70
3.13.3	求解线谱频率参数的快速算法	71
3.14	基于脉冲耦合神经网络的语音特征	73
3.14.1	PCNN 模型结构及其原理	74
3.14.2	PCNN 在图像特征提取中的应用	75
3.14.3	运用 PCNN 从语谱图中提取特征参数	77

3.14.4	语谱图	77
3.14.5	运用 PCNN 提取特征参数	78
3.15	耳蜗滤波器倒谱系数	78
3.15.1	耳蜗滤波器倒谱特征原理框图	78
3.15.2	听觉变换	78
3.15.3	耳蜗滤波器	81
3.15.4	毛细胞窗口	84
3.15.5	非线性响度变换和 DCT	84
3.16	加权组合过零峰值幅度特征参数	84
3.16.1	CZCPA 语音特征参数提取原理	85
3.16.2	差分原理	88
3.16.3	WCZCPA 特征参数提取原理	89
3.17	基于二维根倒谱和 CCBC 的特征参数	90
3.17.1	二维根倒谱(TDRC)	90
3.17.2	CCBC 算法	91
3.17.3	特征参数提取流程	91
3.18	基于改进 LDA 和 PCA 的语音特征参数	92
3.18.1	基于数据驱动线性特征转换的语音特征变换	92
3.18.2	改进的算法及其实现	94
3.19	基于 MUSIC 和感知特性的鲁棒特征参数	95
3.19.1	感知处理	95
3.19.2	PMUSIC-MFCC 特征提取流程	96
3.19.3	实验结果对比	98
3.20	基于小波包变换的鲁棒特征参数	100
3.20.1	小波包分解	100
3.20.2	新参数的算法构想	102
3.20.3	小波包对频带的划分	103
3.20.4	小波函数的选取	105
3.20.5	新参数提取流程	108
3.20.6	实验结果对比	110
3.21	本章小结	112
第 4 章	语音动态特征参数提取	121
4.1	问题的提出	121
4.2	基于 MUSIC 和调制谱滤波的动态特征参数	121

4.2.1	MUSIC 谱估计	122
4.2.2	调制谱原理	124
4.2.3	MMS-MFCC 特征提取流程	126
4.3	基于共振峰曲线的语音信号动态特征参数	130
4.3.1	动态特征提取流程	130
4.3.2	仿真实验	131
4.4	组合特征参数优化选择方法	134
4.4.1	基本正交实验设计简介	135
4.4.2	因子和水平的选择	137
4.4.3	正交表的选用及构造	137
4.4.4	正交实验结果与分析	138
4.4.5	对比实验结果与分析	143
4.5	本章小结	144
第5章	基于小波包变换和 K-L 展开的语音识别方法	148
5.1	问题的提出	148
5.2	小波包变换	148
5.3	特征提取	149
5.4	K-L 展开	150
5.5	仿真实验结果	151
5.6	本章小结	152
第6章	基于小波混沌神经网络的语音识别方法	155
6.1	问题的提出	155
6.2	动态时间规整	155
6.3	小波混沌神经网络模型	155
6.4	WCNN 学习算法	156
6.4.1	小波神经网络部分学习算法	156
6.4.2	混沌神经网络部分学习算法	157
6.5	实验结果及结论	159
6.6	本章小结	160
第7章	基于遗传算法和 Fisher 投影的语音识别方法	162
7.1	问题的提出	162
7.2	遗传算法	162

7.3	基于 Fisher 准则函数的最佳鉴别矢量	163
7.3.1	Fisher 准则函数	163
7.3.2	最佳鉴别矢量基	164
7.4	混沌神经网络模型	164
7.5	实验及结果分析	165
7.6	本章小结	167
第 8 章	语音可视化技术研究	170
8.1	问题的提出	170
8.2	基于语谱图的可视化方法	170
8.3	基于 LLE 和模糊核聚类的可视化方法	171
8.3.1	局部线性嵌入(LLE)方法	172
8.3.2	改进 LLE 方法	173
8.3.3	核方法	173
8.3.4	模糊核聚类算法	175
8.3.5	仿真实验结果与分析	176
8.4	基于集成特征和神经网络的可视化方法	178
8.4.1	音素与音节	179
8.4.2	基音与四声	179
8.4.3	可视化系统结构模块	180
8.4.4	神经网络设计	181
8.4.5	图案信息编码	181
8.4.6	主颜色编码	184
8.4.7	图像位置信息映射	185
8.4.8	图像合成	185
8.4.9	仿真实验及结果分析	185
8.5	基于颜色和图案的汉语声韵母可视化方法	190
8.5.1	可视化系统结构	190
8.5.2	BP 神经网络设计	191
8.5.3	共振峰特征映射主颜色信息	192
8.5.4	谐振强度和音长映射图像长和宽信息	192
8.5.5	组合特征映射图案信息	193
8.5.6	图像合成	193
8.5.7	仿真实验及结果分析	193
8.6	基于神经网络的汉语声韵母可视化方法	198

8.6.1	可视化系统总体结构	198
8.6.2	语音信号获取及预处理	199
8.6.3	语音特征提取及 PCA 降维	199
8.6.4	神经网络设计	200
8.6.5	位置信息映射	200
8.6.6	颜色信息获取	201
8.6.7	图像合成	201
8.6.8	仿真实验及结果分析	201
8.5	本章小结	203
第 9 章	结论与展望	209
9.1	本书主要工作及创新点	209
9.2	进一步研究的展望	211

第1章 绪论

1.1 语音信号研究背景概述

声学是物理学的一个分支学科，而语言声学又是声学的一个分支学科。它主要的研究方向是人的发声器官机理、发声器官的数学模型、听觉器官的特性（如听阈、掩蔽、临界带宽和听力损失等）、听觉器官的数学模型、语音信号的物理特性（如频谱特性、声调特性、相关特性和概率分布等）、语音的清晰度和可懂度等。当今通信和广播的发展非常迅速，而语言通信和语言广播仍然是最重要的部分，语言声学则是这些技术科学的基础。

语言声学的发展和电子学、计算机科学有着非常密切的关系。在它发展的过程中有过几次飞跃，第一次飞跃是1907年电子管的发明和1920年无线电广播的出现。因为有了电子管放大器，很微弱的声音也可以放大，而且可以定量测量，从而使电声学和语言声学的一些研究成果扩展到通信和广播部门。第二次飞跃应该是在20世纪70年代初，由于电子计算机和数字信号处理的发展，人们发现：声音信号特别是语音信号，可以通过模数转换器(A/D)采样和量化，将其转换为数字信号，然后送进计算机，这样就可以用数字计算方法，对语音信号进行处理和加工。例如，频谱分析可以用傅里叶变换或快速傅里叶变换实现，数字滤波器可以用差分方程实现。在这个基础上，逐渐形成了一门新学科——语音信号处理。它的发展很快，在通信、自动控制等领域，解决了很多用传统方法难以解决的问题，在信息科学中占有很重要的地位。

在现代信息社会中，小至人们的日常生活，大到国家大事、世界新闻、社会舆论和各种重要会议，都离不开语言和文字。近年来，普通电话、移动电话和互联网已经普及到家庭。在这些先进的工具中，语音信号处理中的语音编码和语音合成就有很大贡献。再进一步，可以预料到的口呼打字机（又称听写机，它能把语音转换为文字）、语音翻译机（例如输入为汉语，输出为英语，或者相反），已经不是梦想，而是提到日程上的研究工作了。人们早就希望用语音指挥机器，机器的执行情况也能用语音回答，这在某些领域已经部分地实现

了。目前，计算机芯片的集成度和运算能力，每 18 个月就提高一倍，而成本又不断降低，因此，它已经广泛地应用在社会生产和生活的各个方面。然而计算机接收信息的外围设备和主机相比，要逊色得多，能说能听的计算机还不能普遍使用，也就是说：语音识别、语音可视化、语音理解和语音合成等课题，还有很多理论问题和技术问题没有解决，需要继续深入研究。

科学家们深入研究后认为，要解决人一机语音对话这样的难题，做出真正实用的语音机器，必须开展跨学科的研究，如声学、语言学、语音学、生理学、数字信号处理、人工智能和计算机科学等。要真正赋予微电脑以语言功能，必须彻底了解语言是如何产生、感知，以及人类的语言通信是如何进行的。图 1.1 给出了从语言产生到语音感知全过程中的几个重要环节，从图中可以看到，要使这个问题得到满意的解决，需要深入研究人类发声器官和听觉器官机理，建立能反映客观真实情况的物理模型和数学模型。

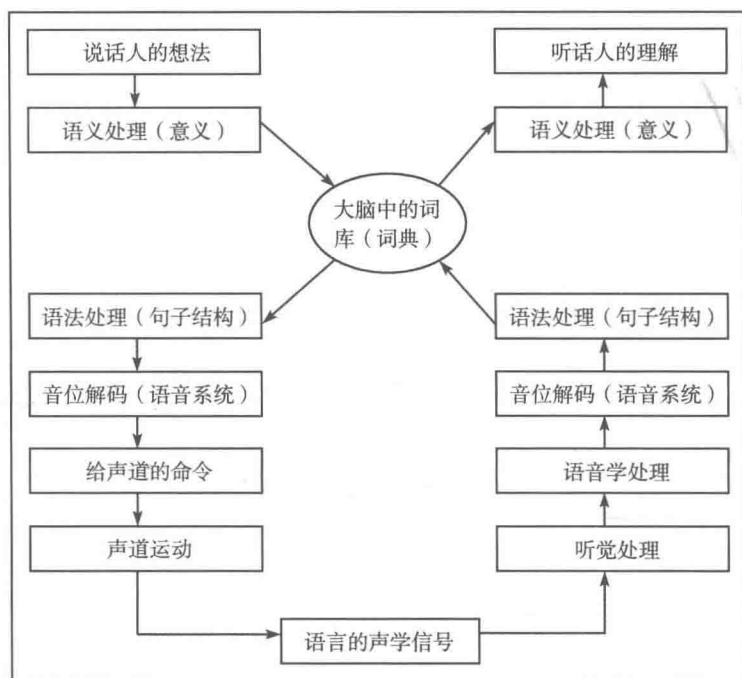


图 1.1 人类语音通信的过程

1.2 国内外研究现状

1.2.1 语音识别技术研究

语音是人类之间最有效、最方便的通信方式。语音识别技术(又称自动语音识别, Automatic Speech Recognition, ASR)是从20世纪50年代才开始出现的一门新兴的计算机智能技术。其研究目的是让机器“听懂”人类的语言,即可以使机器通过识别和理解过程把语音信号转变为相应的文本或指令,或者针对人类的语音要求及询问,能够理解其意图并做出正确的响应,从而实现人机自然语言通信。

语音识别最早期的研究始于20世纪50年代,以1952年美国贝尔实验室发表的关于特定人、小词汇量孤立数字识别系统的研究论文为起点。随后美国RCA研究所和MIT Lincoln实验室做了进一步的深入研究,分别于1956年和1959年成功研发了10音节特定人语音识别系统和10元音非特定人语音识别系统。这个时期对语音识别的研究还处于初始阶段,主要依靠不同元音频谱间的差别来对不同发音进行区分。直到20世纪60—70年代,随着数字信号处理领域的理论算法成熟和计算机产业的迅速发展,语音识别才作为一个重要的研究课题而展开,并逐步取得一系列实质性的进展。这个时期的研究以孤立词语音识别为主,基于模板匹配原理,线性预测(Linear Prediction, LP)技术和动态时间规整(Dynamic Time Warping, DTW)算法被成功引入到语音信号处理中,有效地解决了说话人语速不均匀造成时间伸缩变化的影响,掀起了语音识别研究的热潮。与此同时,对非特定人语音识别的数据分析聚类方法等一系列重要的技术对之后的语音识别研究也产生了深远的影响。

进入20世纪80年代以后,语音识别研究进一步走向深入,随着词汇量的逐渐增多,研究重点由孤立词的语音识别转为连接词的语音识别,各种连接词语音识别算法被开发出来,用于连接词识别的分层构筑技术(Level Building)得到发展。由于很难对自然的连续语音进行分割,语音单元间的协同发音现象十分普遍,因此,基于模板匹配结构的识别系统不再适用。这个时期语音识别研究的重点转为基于统计模型(Statistical Language Modeling, SLM)的识别技术,人们研究从微观转向宏观,不再刻意追求细化语音特征,而是更多从整体平均(统计)的角度来建立最佳的语音识别系统,在统计模型的框架下寻找令模型参数最大化的词汇作为识别结果,隐马尔可夫(Hidden Markov Models, HMM)模型和人工神经网络(Artificial Neural Networks, ANN)被应用到语音识别领域,

在语音识别中获得极大的成功,开创了语音识别的新时代。20世纪80年代美国在语音识别方面进行的一些重大研究项目都采取以HMM为基本框架的统计途径,其中以1988年美国CMU大学用VQ/HMM方法实现的世界第一个高性能、非特定人、大词汇量(997词)连续语音识别系统SPHINX为代表。此外,AT&T公司、Bell实验室以L. R. Rabiner为首的科研集团在连接数字识别和语声响应(voice response)等方面的工作、IBM公司以F. Jelinek为首的研究组在语音打字机方面所做的工作(Tangora系统),以及美国国防部的高级研究规划局(American Research Projects Agency, ARPA)重新制定的新五年计划——DARPA计划——都在很大程度上推动了语音识别技术的进步,其后,连续语音识别技术获得了长足的发展。

20世纪90年代后,语音识别在细化模型设计、参数提取和优化以及系统的自适应等方面取得一系列关键性的进展,使得语音识别技术进一步成熟,伴随着多媒体时代的来临,语音识别技术由实验室理论仿真逐步走向市场实用,进入了商品化开发阶段。其中以IBM的ViaVoice听写机,AT&T的电话系统,剑桥大学的HTK系统、OGI系统、DARGON系统和Microsoft的Whisper系统为代表。当今,基于HMM和ANN相结合的方法受到了广泛重视。而一些模式识别、机器学习方面的新技术也被应用到语音识别中,如支持向量机(Support Vector Machine)技术和进化计算(Evolutionary Computation)技术等^[1-4]。

我国语音识别研究工作起步于20世纪50年代,但直到70年代才开始迅速发展。中国科学院、清华大学、北京大学等多家研究单位在从事汉语语音识别系统的开发,目前,对大词汇量连续语音识别系统的研究已经接近国外最高水平。在我国的“八五”计划和“863”计划中,汉语语音识别的研究得到了大力支持,国家863《智能计算机主题》专家组专门为语音识别研究立项,同时,由于中国在国际上地位与日俱增,以及在经济和市场方面所处的重要地位,汉语语音识别也越来越被国外研究机构和公司重视,IBM、Microsoft、APPLE、Motorola、Intel、L&H等公司都在国内设立研究机构,相继投入到汉语语音识别系统的开发中,强有力地推动了汉语语音识别研究的发展^[5]。

尽管如此,距离真正的人机自由交流的境界还很遥远。现在已有的商用系统都存在着一些问题,比如对于噪声环境下的语音识别率和稳健性等都不尽如人意。不可否认,语音识别技术还有一段路需要走,要做到真正成功的商业化,它还需要在很多方面取得突破性进展,未来语音识别的发展将会更加迅速,它将逐渐深入到我们生活的方方面面。

1.2.2 语音可视化技术研究

据1987年国家统计局对全国残疾人抽样调查公布的数字,由于听力及智