摘 要

语音识别的研究工作始于上个世纪 50 年代，至今已经形成了完整的理论体系，目前语音识别的研究也已经进入了商品化阶段，基础性理论相当完善，各种各样的产品也相继涌现。然而语音识别在实现过程中通常涉及多种因素，需要同时考虑，并且它作为一门交叉学科，涉及到了信号处理、模式识别、人工智能、计算机科学、语言学和认知科学等众多学科，所以语音识别距离理想目标仍有很大距离，相关的技术难关还有待克服。

本文对语音识别的主要过程进行了详细的介绍。语音识别首先对输入的语音信号必须进行预处理，以保证系统获得一个比较理想的处理对象。在语音的特征参数提取阶段，文中介绍了在实际应用中常用到的特征参数：线性预测倒谱参数(LPCC)、Mel 频率倒谱参数(MFCC)等。在识别阶段，介绍了基于矢量量化的识别技术以及动态时间归整的识别技术(DTW)。在此基础上，引入了蚁群算法的基本原理。

蚁群算法是最新发展的一种模拟昆虫王国中蚂蚁群体智能行为的仿生优化算法，它具有较强的鲁棒性、优良的分布式计算机制、易于与其他方法相结合等优点。蚁群算法作为一种新的用于解决复杂优化问题的全局搜索方法，已经成功应用于求解 TSP 问题、调度问题、指派问题等，显示出了蚁群算法在处理复杂优化问题方面的优越性。

本文利用蚁群算法优化机制，结合传统的 DTW 算法，提出了一种新的基于蚁群算法的动态时间规划算法来搜索语音信号特征参数序列之间匹配的全局最优路径，进而以此衡量语音信号之间的相似度，从而使系统的识别效果有了进一步的提高。

文中最后对新的语音识别系统各模块进行了仿真测试，给出了仿真计算结果。实验结果表明，采用基于蚁群算法的语音识别系统识别效果要好于采用传统 DTW 算法的语音识别系统。

关键词：语音识别，端点检测，蚁群算法，DTW

Abstract

The speech recognition which has been researched since the 1950s, has developed to an integrated theory and been commoditized with perfect basic theory and lots of products successively emerging. However, the practice of speech recognition is related to various factors, which must be considered simultaneously in the process. As a cross-discipline, it also has everything to do with many subjects, such as signal processing, pattern recognition, artificial intelligent, computer science, linguistics and epistemic science. Therefore, there are still many associated technological difficulties to be conquered and the current speech recognition is still far from the final target.

The main process of speech recognition is analyzed and investigated thoroughly. First, the input of speech signals must be pre-processed in the system in advance, so that the object for the system to process is comparatively ideal. Secondly, the frequently-used characteristic parameters, such as LPCC and MFCC, are introduced in detail when coming to abstracting characteristics while some key techniques including VQ and DTW are analyzed in the recognition step. Then, the basic principles of ant colony algorithm are introduced.

Ant colony algorithm which is one of the algorithms latest developed, is a bionic optimization algorithm by simulating the intelligence of ants swarm in insect kingdom. As a new algorithm used to solve complex optimization problems of global search method, the ant colony algorithm with its robustness, good distributed computing mechanism and easy-combination with other methods has been successfully applied into TSP, scheduling and assignment problems, showing many advantages in dealing with the complex optimization problems.

By combining ant colony algorithm optimization mechanism with the traditional DTW algorithm, a new dynamic time programming algorithm based on the ant colony algorithm is proposed, which is used to search the speech signals characteristic parameters sequences for the global optimal path, by which the similarity between the speech signals is measured. Thus, the recognition result of the

system has been further improved.

The new speech recognition system is tested by simulating every single module and evaluated with the result figures shown in the final part. The experimental results illustrate that the speech recognition system based on ant colony algorithm has better performance than that based on traditional DTW algorithm.

Keyword: Speech recognition, endpoint detection, ant colony algorithm, DTW

独 创 性 声 明

本人声明，所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含基他人已经发表或撰写过的研究成果，也不包含为获得武汉理工大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名：肖宜　　日期：2008年5月19日

关于论文使用授权的说明

本人完全了解武汉理工大学有关保留、使用学位论文的规定，即学校有权力保留、送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名：肖宜　导师签名：[illegible]　日期：2008年5月19日

第1章 绪论

1.1 研究背景与意义

语言是人类创造的，也是人类区别于地球上其它生物的本质特征之一[1]，更是人类最重要的交流工具，有着自然、方便、准确性高等特点。随着计算机和人工智能机器的广泛应用，人们发现：人与机器最方便、最直接的沟通方式是语言通信。让机器听懂人类所说的话，明白人所表达的意思，并且根据说话者的意思而做出相应的动作，这就是语音识别技术。

语音识别就是指智能机器自动识别语音的技术[1]，有广义和狭义之分。广义上的语音识别技术是指识别出语音信号中“感兴趣的内容”其中包括：识别说话人的内容、说话人的身份、说话人的语言等。而狭义上语音识别技术是指准确的识别出语音信号所表达的意思，准确的理解语音信号所表达的含义。在计算机普及的今天让计算机听懂人的语言是人类所向往的事，对计算机直接用语言发号施令，解放我们的双手就显得特别重要了。世界上各大 IT 的著名公司如：Philips、IBM、Intel 等都投入巨大的财力、精力对语音识别进行研究。微软总裁盖茨曾经就说过：“我们将在这几十年中，克服语音识别的障碍，下一代的系统操作软件及应用程序的用户界面将抛弃键盘与鼠标，代以真正意义上的人机对话[2]。”

蚁群算法作为一种新的用于解决复杂优化问题的全局搜索方法，已经成功应用于求解 TSP 问题、调度问题、指派问题等，显示出了蚁群算法在处理复杂优化问题[3,4]方面的优越性。蚁群算法最早由意大利学者 Dorigo 等人[5,6]在 20 世纪 90 年代初首先提出来的，它是受自然界中的蚂蚁集体行为的启发而提出的，算法具有分布式计算、信息正反馈和启发式搜索的特征，本质上是进化算法中的一种新型随机性优化算法。

在基于模板匹配的语音识别系统中，DTW(dynamic time warping)算法被广泛应用。但 DTW 算法是一种局部最优算法，其每一步搜索都是根据局部优化的判断进行的，因此这个时间规整路径不一定达到全局最优。利用蚁群算法优化机制，结合传统的 DTW 算法，可以为基于模板匹配的语音识别系统做出一些有意义的研究工作。

1.2 国内外研究现状

语音识别的研究工作起源于20世纪50年代，1952年AT&T Bell实验室的Davis等人实现了第一个可识别十个英文数字的语音识别系统——Audry系统，这个系统主要依赖每个数字辅音部分的频谱进行识别。1956年RCA实验室的Olson等人也独立地研制出10个单音节部分的识别系统，系统采用从带通滤波器组获得的频谱参数作为语音的特征。1959年Fry和Denes等人尝试构建音素识别器来识别4个元音和9个辅音，并采用频谱分析和模式匹配来进行识别策略。与此同时MIT林肯实验室的Forgie等人研究了10个元音的识别，并采用了声道时变估计技术。

20世纪60年代，1960年G. Fant在其论作《语音产生的声学原理》中提出了语音产生的声源——滤波器模型，为语音信号参数的处理提供了理论基础。随后计算机的应用推动了语音识别的发展。这时期推出的两大关键技术：动态规划（DP，Dynamic Programming）和线性预测分析技术（LP，Linear Prediction），对语音识别发展意义深远。值得一提的是60年代中期，美国斯坦福大学的Reddy就开始尝试用动态跟踪音素的方法来进行连续语音识别，开展了卓有成效的工作。

20世纪70年代，语音识别进入一个新的里程碑，日本学者Sakoe给出了动态时间弯折算法（DTW，Dynamic Time Warping），并在实际中得到应用，实现了基于特定人孤立词语音识别系统。DTW是一种模式匹配和模型训练技术，它应用动态规划方法成功解决了语音信号特征参数序列比较时时长不等的难题，在孤立词语音识别中获得了良好性能。Itakura基于语音编码中广泛使用的线性预测编码(LPC，Linear Predictive Coding)技术，通过定义基于LPC频谱参数的合适的距离测度，成功地将其扩展到语音识别中。

20世纪80年代，Linda、Buzo、Gray等人解决了矢量量化(VQ，Vector Quantization)码本生成的方法，并将矢量量化技术成功地应用到语音编码中。随后语音识别研究进一步走向深入，出现了大量连续语音的识别算法。典型代表为Bell实验室推出的分层构造（LB，Level Building）技术。到了1988年美国的卡内基-梅隆大学运用矢量量化(VQ，Vector Quantization)和隐马尔可夫（HMM，Hidden Markov Models）技术开发了针对非特定人连续语音的SPHINX系统，在语音识别方面取得了巨大的成功,这是世界上第一个高性能的非特定

人、大词汇量、连续语音识别系统。

到了 80 年代后期，人工神经网络(ANN，Artificial Neural Network)技术用于语音识别也开始广泛开展，大部分采用基于反向传播算法的多层感知网络。ANN 具有区分复杂的分类边界能力，所以有助于模式的区分。

进入 20 世纪 90 年代以后，语音识别从实验室走向实用。许多发达国家如美国、日本、韩国以及 IBM、AT&T、L&H 等著名公司都为语音识别系统的实用化开发研究投以巨资。AT&T 开发了能识别英文发音卡号的信用卡语音系统。IBM 公司率先推出 Via Voice 大词汇量非特定人汉语连续语音识别系统，Microsoft 公司也开发了中文识别引擎，两者代表了当时汉语识别的最高水平。

从 90 年代末开始，一些大规模的语音识别系统在实际中开始广泛应用。1996 年 9 月，Charles Schwab 开通了首个大规模商用语音识别应用系统：股票报价系统。该系统有效地提高了服务质量和客户满意度，并减少了呼叫中心的费用。随后 Schwab 又开通了语音股票交易系统。美国主要电信运营商 Sprint 的 PCS 部门自 2000 年来为客户开通了语音驱动系统，提供客户服务、语音拨号、查号和更改地址等业务，2001 年 9 月开通的可以以自然方式对话的咨询系统，更实现了以自然、开放的询问方式实时获得所需要的信息。

我国语音识别研究工作相对国际水平起步较晚，1986 年我国高科技发展计划(863 计划)启动，语音识别作为智能计算机系统研究的一个重要组成部分而被专门列为研究课题，从此我们开始有组织地进行语音识别技术的研究。经过二十余年的发展，理论上也逐渐成熟，越来越多的大学和研究所加入到语音识别研究中来。清华大学的王作英教授提出了一个基于段长分布的非齐次隐马尔可夫模型。以此理论为指导所设计的语音识别听写机系统在 1998 年的全国语音识别系统 863 评测中取得冠军，从而显示了这一新模型的生命力和在这一研究领域内的领先水平。2002 年中科院自动化所推出的面向不同计算平台和应用的中文语音系列产品——Pattek ASR，表明我国“863”高技术领域的又一重量级核心技术破土而出，也是我国首次拥有完全自主知识产权并形成产品化的语音识别技术。

国内外众多媒体和专家将语音识别技术评为 21 世纪前十年将对人类生活方式产生重大影响的十大科技进展之一。比尔·盖茨预测：“未来十年语音技术将成为主流。”中国互联网络中心也预测：“未来五年，中文语音技术领域将会有 1300 亿元的市场容量。”

1.3 语音识别系统的分类

一个复杂的语音识别系统，根据服务对象、词汇量大小、工作环境、发音方式、任务性质等诸多因素的不同，可以分为以下几类[41]：

（1）按发音方式分类

按发音方式语音识别系统可分为孤立词语音识别系统、连接词语音识别系统和连续语音识别系统。

孤立词语音识别系统指人在发音时，以单个词的发音方式向语音识别系统输入语音，词与词之间要有足够的时间间隙，以便系统能够检测到始末点。采用这种方式的语音识别系统已经有了较为成熟的算法，实现起来较为容易。连接词语音识别系统指以词或词组为发音单位向系统输入语音。与孤立词发音相比，这种发音方式比较自然，且输入效率也比较高。中小词汇量连接词语音识别系统的识别率目前可以做得很高，并达到了实用水平。连续语音识别系统指在输入语音时，完全按照人的最自然的说话方式输入。这种系统是最方便的输入系统，但是，实现起来也是最复杂和最困难的。

（2）按应用对象分类

按应用对象语音识别系统可分为特定人和非特定人识别系统。特定人的语音识别系统，对于每一个使用者都必须建立专用的参考模板库。非特定人语音识别的原则是事先用许多人(通常 30-40 人)的语音样本训练系统，使用者无论是否参加过采样训练都可以只用一套参考模板，使用该系统进行语音识别。

这两类系统的应用对象大不相同，为了达到良好的识别效果，其系统结构、特征参数选择以及识别方法都可能有极大的差别。对于非特定人的语音识别系统来说，由于要考虑各种复杂因素，实现起来要比特定人的语音识别系统困难得多。

（3）按识别词汇量的大小分类

按词汇量的大小可分为小词汇量识别系统、中等词汇量识别系统、大词汇量识别系统和无限词汇量识别系统。随着词汇数目的增加，潜在的词间相似性会增加，系统的搜索运算开销及存储开销相应增加，识别系统的难度一般也会增加。当系统所能识别的词汇量越大时，实现起来就越困难[9,41]。

1.4 本文研究的主要内容

目前，语音识别系统大多采用模式匹配的原理。本文分析完整的语音识别系统的系统结构和系统的各个模块，利用蚁群算法优化机制，结合传统的 DTW 算法，设计出一种蚁群动态时间规划算法。基于本课题的研究内容和主要工作，本文的结构如下：

第一章主要介绍了本课题的背景、目的和意义，同时介绍了语音识别的国内外研究现状。

第二章分析完整的语音识别系统的系统结构与系统的各个模块，讨论了语音的预处理、端点检测等识别技术，并讨论了经典的语音特征参数，即线性预测倒谱系数（LPCC，Linear Prediction Cepstrum Coefficient）和 Mel 频率倒谱系数（MFCC，Mel Frequency Cepstrum Coefficient）以及经典模版匹配识别技术(DTW，Dynamic Time Warping)，为后来的蚁群动态时间规划算法的引入打下基础。

第三章详细介绍了经典蚁群算法的的原理及特点，结合传统的 DTW 算法，提出基于蚁群算法的动态时间规划算法，详细介绍了基于蚁群算法的动态时间规划算法的基本原理，蚂蚁构造路径，信息素更新机制。

第四章采用蚁群动态时间规划算法实现语音识别系统，对语音信号的预处理、端点检测以及特征提取进行了仿真实现，并与 DTW 算法对比做了动态时间规划的仿真试验。

第五章进行了系统的总结以及对进一步工作的展望。

第 2 章　语音识别系统的系统结构分析

2.1　语音识别系统总体结构

从总体上看，语音识别处理过程可以由一个框架来表示。其结构如图 2-1 所示。从这个总体结构可以看出：语音识别对输入的语音信号首先要进行预处理，对信号进行适当放大和增益控制，并进行预加重和端点检测。然后进行数字化，将模拟信号转化为数字信号以便用计算机来处理，接着进行特征提取，用反映语音信号特点的若干特征参数来代表话音。对特征参数的要求是：(1)提取的特征参数能有效地代表语音特征，具有很好的区分性；(2)各阶参数之间有良好的独立性；(3)特征参数要计算方便，最好有高效的计算方法，以保证语音识别的实时实现。常用的特征包括短时平均能量或幅度、短时平均过零率、短时自相关函数、线性预测系数、短时傅里叶变换和倒谱等。

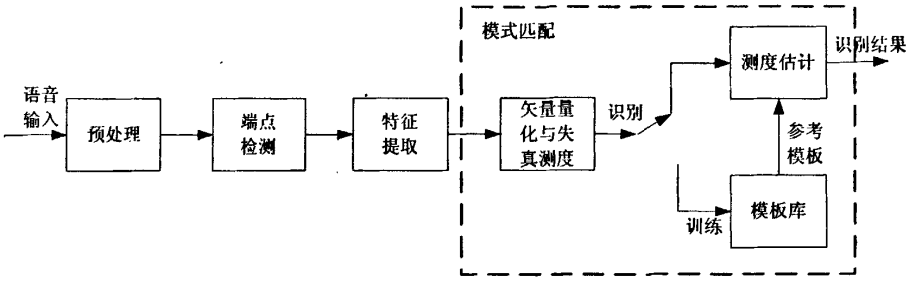

图 2-1 语音识别系统的原理框图

语音识别技术分为两个阶段：训练阶段和识别阶段。在训练阶段，对用特征参数形式表示的语音信号进行相应的处理，获得表示识别基本单元共性特点的标准数据，以此构成参考模板，将所有能识别的基本单元的参考模板结合在一起，形成参考模式库；在识别阶段，将待识别的语音经特征提取后逐一与参考模式库中的各个模板按某种原则进行比较，找出最相像的参考模板所对应的发音，即为识别结果[7]。

2.2 语音信号的预处理与端点检测

2.2.1 语音信号的采样与预加重

根据 Nyqusti 采样定理，如果模拟信号的频谱带宽是有限的(例如不包含高于 f_m 的频率成分)，那么用不小于 $2f_m$ 的取样领率进行取样，则能从取样信号中恢复出原模拟信号[8]。就语音信号而言，浊音语音的频谱一般在 4kHZ 以上便迅速下降，而清音语音信号的频谱在 4kHz 以上频段反而呈上升趋势，甚至超过了 8kHz，以后仍然没有明显下降的趋势[8]。因此，为了精确表示语音信号，一般认为必须保留 10kHz 以下的所有频谱成分，这意味着采样频率应当等于或大于 20kHz。但是在许多实际应用中并不需要采用这么高的取样频率，实验表明对语音清晰度和可懂度有明显影响的成分，最高频率约为 5.7kHz。例如 ITU(International Teleeommunieation Union，国际电信联盟)在 G.729 中提出的语音编解码系统采样频率为 8kHz，只利用了 3.4kHZ 以内的语音信号分量[9,10]，虽然这样的采样频率对语音清晰度是有损害的，但受损失的只是少数辅音，而语音信号本身的冗余度又比较大，少数辅音清晰度下降并不明显影响语句的可懂度。因此语音识别时常用的采样频率为 8kHz、10 kHz 或 16 kHz。本课题采用了 8kHz 和 16kHz 两种采样频率进行试验。

语音信号在采样之前要进行预滤波处理。预滤波的目的是：(1)抑制输入信号各频率分量中频率超过 $f_s/2$ 的所有分量(f_s 为采样频率)，以防止混叠干扰；(2)抑制 50Hz 的电源干扰。进行预预滤波处理后，再采用合适的采样频率进行采样。目前，较好的声卡通常都带有带通滤波器。

由于语音信号的平均功率谱受到声门激励和口鼻辐射的影响，语音信号从嘴唇辐射后有 $6dB/Oct$ (倍频程)的衰减。因此，在对语音信号进行分析之前，要对语音信号的高频部分加以提升，利用在处理前提升声音中高频达到减小噪声的效果，使得语音信号的频谱变得平坦，压缩信号器的动态范围，提高信噪比。提升的方法有两种：其一是用模拟电路实现；其二是用数字电路实现。采用数字电路实现 $6dB/Oct$ 预加重的数字滤波器的形式为：

$$y(n) = x(n) - \alpha x(n-1) \tag{2-1}$$

其中：$x(n)$ 为原始语音序列；$y(n)$ 为预加重后的序列；α 为预加重系数。通常，α 的值取 0.9~1.0 之间的数，通常取 0.98 或者 0.97[11]。本课题采用 α=0.98

进行语音的预加重。

2.2.2 语音信号的加窗

语音信号是一种典型的非平稳信号，其特性是随时间变化的。但是，语音的形成过程是与发音器官的运动密切相关的，这种物理运动比起声音振动速度来讲要缓慢得多，因此语音信号常常可假定为短时平稳的，即在 10~20ms 这样的时间段内，其频谱特性和某些物理特征参量可近似地看作是不变的。这样，就可以采用平稳过程的分析处理方法来处理了。由此导出了各种“短时”处理方法，以后讨论的各种语音特征参数的提取都基于这个假定。这种依赖于时间处理的基本方法，是将语音信号分隔为一些短段(帧)再加以处理。这些帧就好像是来自一个具有固定特性的持续语音片段一样，一般都按要求重复(常是周期的)，对每帧语音进行处理就等效于对固定特性的持续语音进行处理。短段之间彼此经常有一些重叠，对每一帧的处理结果是一个数或是一组数[12]。经过处理后将从原始语音序列产生一个新的依赖于时间的序列，被用于描述语音信号的特征。

设原始语音信号采样序列为 $x(m)$，将其分成短段等效于乘以幅度为 1 的移动窗 $x(n-m)$。当移动窗幅度不是 1 而是按一定函数取值时，所分成的短段语音的各个取样值将受到一定程度的加权。

对语音信号的各个短段进行处理，实际上就是对各个短段进行某种变换或施以某种运算，其一般式为：

$$Q_n = \sum_{m=-\infty}^{+\infty} T[x(m)]w(n-m) \tag{2-2}$$

其中 $T[*]$ 表示某种变换，它可以是线性的也可以是非线性的，$x(m)$ 为输入语音信号序列。Q_n 是所有各段经过处理后得到的一个时间序列。

对语音信号加窗时，用的最多的三种窗函数是矩形窗、汉明窗(Hamming)、汉宁窗(Hanning)，其定义分别为：

（1）矩形窗：$$w(n)=\begin{cases}1; & 0\le n\le L-1\\ 0; & other\end{cases} \tag{2-3}$$

（2）汉明窗：$w(n)=\begin{cases}0.54-0.46\cos\left(\dfrac{2\pi n}{L-1}\right); & 0\le n\le L-1\\ 0; & other\end{cases}$ (2-4)

（3）汉宁窗：$w(n)=\begin{cases}0.5\left[1-\cos\left(\dfrac{2\pi n}{L-1}\right)\right]; & 0\le n\le L-1\\ 0; & other\end{cases}$ (2-5)

其中 L 为窗长。窗函数越宽，对信号的平滑作用越显著，窗函数过窄，对信号平滑作用越不明显。对波形乘以窗函数，相当于在频谱范围内，对信号的频谱进行窗函数的付里叶变换的卷积，或者是进行加权移动的平均。一般希望窗函数只有以下的性质：一是频率分辨高，即主瓣狭窄、尖锐；二是频谱泄漏少，侧瓣衰减大。由于汉明窗在频率范围中的分辨率较高，而且侧辨的衰减大于 43dB，具有频谱泄漏少的优点，所以在本课题的语音识别系统中，采用 Hamming 窗作为窗函数[11]。

2.2.3 语音信号的端点检测

语音信号起止点的判别是任何一个语音识别系统必不可少的组成部分。因为只有准确的找出语音段的起始点和终止点，才有可能使采集到的数据是真正要分析的语音信号，这样不但减少了数据量、运算量和处理时间，同时也有利于系统识别率的改善[12]。因此端点作为语音分割的重要特征，在很大程度上影响语音识别系统的性能，如何在噪声环境下设计一个鲁棒的端点检测算法是一个非常棘手的问题。常用的端点检测方法有以下几种。

(1)短时平均能量

设 $S(n)$ 为加窗语音信号，第 t 帧语音的短时平均能量为：

$$Eng(t)=\frac{1}{N}\sum_{n=0}^{N-1}|S_t^2(n)| \tag{2-6}$$

$$Or \qquad Eng(t)=\frac{1}{N}\sum_{n=0}^{N-1}|S_t(n)| \tag{2-7}$$

其中 N 为分析窗宽度，$S_t(n)$ 为第 t 帧语音信号中的第 n 个点的信号样值。上面两式原理是相同的，但后式有利于区别小取样值和大取样值，不因前式取平方造成很大差异[13,14]。

短时平均能量是时域特征参数。把它用于模型参数时，应进行归一化处理，本文语音识别系统中取其对数值后使用，使计算和识别结果均取得了较好的效率和结果。

(2)短时过零率

短时过零率 ZCR(Zero-Crossing-Rate)为：

$$Z_n = \sum_{m=-\infty}^{+\infty} Sgn[x(m)] - Sgn[x(m-1)] \times W(n-m) \tag{2-8}$$

其中：

$$\begin{cases} Sgn[x(n)] = 1 & x(n) > NoiseMax(NoiseMax为噪声上限) \\ Sgn[x(n)] = -1 & x(n) < NoiseMax(NoiseMax为噪声下限) \\ Sgn[x(n)] = 0 & otherwise \end{cases} \tag{2-9}$$

$$\begin{cases} W(n) = \dfrac{1}{2N} & 0 \le n \le N-1(N为一阵声音的长度) \\ W(n) = 0 & otherwise \end{cases} \tag{2-10}$$

有噪声的情况下，单纯用短时能量或者短时过零率不能准确检测出语音信号。本课题采用短时能量和短时过零率相结合的方法，利用短时能量和短时过零率两个门限来确定语音信号的起点和终点，目的是从采集到的语音信号中分离出真正的语音信号作为系统处理的对象。

2.3 特征参数的提取

语音识别的首要步骤是特征提取，有时也称为前端处理，与之相关的内容则是特征间的距离度量。所谓特征提取，即对不同的语音寻找其内在特征，由此来判别出未知语音，所以每个语音识别系统都必须进行特征提取。特征的选择对识别效果至关重要，选择的标准应体现对异音字之间的距离尽可能大，而同音字之间的距离应尽可能小。若以前者距离与后者距离之比为优化准则确定目标量，则应是该量最大。同时，还要考虑特征参数的计算量，应在保持高识别率的情况下，尽可能减少特征维数，以减小存储要求并利于实时实现[15,16]。

语音的特征参数多种多样，在实际应用中，可以根据需要选择不同的语音参数或几种参数的组合。在语音识别中经常用到的特征参数有 LPC 倒谱参数(LPCC)和 Mel 频率倒谱参数(MFCC)等。

2.3.1 线性预测系数

线性预测（Linear Prediction）基本思想是由于语音信号样点之间存在相关性，所以可以用过去的样点值来预测现在或未来的样点值，即一个语音的抽样能够用过去若干个语音抽样的线性组合来逼近，通过使实际语音信号抽样值和线性预测抽样值之间的误差在均方准则下达到最小值来求解预测系数，而这组预测系数就反映了语音信号的特征，故可以用这组语音特征参数进行语音识别或语音合成等。

（1）线性预测的基本原理

若一个随机过程用一个 p 阶的全极点系统受白噪声激励产生的输出来模拟，设这个系统的传递函数为：

$$H(z)=S(z)/U(z)=\frac{G}{1-\sum_{k=1}^{p}a_k s^{-k}} \tag{2-11}$$

其中 G 为增益常数，$S(z)$ 和 $U(z)$ 分别为输出信号 $s(n)$ 和输入信号 $u(n)$ 的 Z 变换，那么 $s(n)$ 和 $u(n)$ 的关系可以表示为差分方程：

$$s(n)=\sum_{k=1}^{p}a_k s(n-k)+Gu(n) \tag{2-12}$$

观察上式，可以将与 $\{a_k\}$ 有关的部分理解为用信号的前 p 个样本来预测当前样本，即定义预测器：

$$\hat{s}(n)=\sum_{k=1}^{p}a_k s(n-k) \tag{2-13}$$

由于预测系数 $\{a_k\}$ 在预测过程中看作常数，所以它是一种线性预测器，这种预测器最早用于语音编码，因此称为线性预测编码（Linear Predictive Coding, LPC），该预测器的系统函数为：

$$H(s)=\sum_{k=1}^{p}a_k s^{-k} \tag{2-14}$$

可见，如果信号 $s(n)$ 符合公式(2-11)所描述的模型假定，那么用公式(2-13)作为线性预测器对信号 $s(n)$ 的预测，其误差应为：

$$e(n)=Gu(n) \tag{2-15}$$

但是，实际信号不是精确地符合这个假定，因此实际的预测误差应为：

$$e(n)=s(n)-\hat{s}(n)=s(n)-\sum_{k=1}^{p}a_k s(n-k) \tag{2-16}$$

上式表明预测误差序列是信号 $s(n)$ 通过一个具有如下系统函数产生的输出：

$$A(s)=1-\sum_{k=1}^{p}a_k s^{-k} \tag{2-17}$$

比较上式与式(2-11)可知，预测误差滤波器 $A(z)$ 是系统 $H(z)$ 的逆滤波器，即：

$$A(s)=G/H(z) \tag{2-18}$$

由于给定的只有信号 $s(n)$ 和一个未知的模型公式(2-11)，要想这个模型尽可能精确地描述信号 $s(n)$，应使公式(2-16)所得到的预测误差在某一短时的总能量尽可能小，并在此准则下求出最佳预测系数 $\{a_k\}$。为此定义短时平均预测误差能量：

$$\begin{aligned} E_n &= \sum_j e_n^2(j) \\ &= \sum_j \left[s_n(j)-\hat{s_n}(j) \right]^2 \\ &= \sum_j \left[s_n(j)-\sum_{k=1}^{p}a_k s_n(j-k) \right]^2 \end{aligned} \tag{2-19}$$

其中 $s_n(j)$ 是在抽样点 n 附近选择的一个语音帧，即：

$$s_n(j)=s(n+j) \tag{2-20}$$

使公式 2-19 中 E_n 为最小的 $\{a_k\}$ 必定满足

$$\frac{\partial E_n}{\partial a_i}=0(i=1,2,\cdots,p) \tag{2-21}$$

由此便得到以 $\{a_k\}$ 为变量的线性方程组：

$$\sum_{k=1}^{p}a_k\phi_n(i,k)=\phi_n(i,0) \qquad i=1,2,\cdots,p \tag{2-22}$$

其中：

$$\phi_n(i,k)=\sum_j s_n(j-i)s_n(j-k) \tag{2-23}$$

该线性方程组通常有唯一解，一旦解出其中的变量$\{a_k\}$，便可得到一种最小预测误差能量计算公式：

$$\begin{aligned}\hat{E}_n &= \sum_j s_n^2(j) - \sum_{k=1}^{p} a_k \sum_j s_n(j) \times s_n(j-k) \\ &= \phi_n(0,0) - \sum_{k=1}^{p} a_k \phi_n(0,k)\end{aligned} \tag{2-24}$$

由公式(2-16)计算出的最小预测误差序列$e(n)$称为预测残差序列。$\hat{E}_n$就是预测残差能量。

对于增益因子G，因为其在短时内为一个常数。根据公式(2-15)和(2-16)，有：

$$\hat{E}_n = \sum_j e^2(j) = \hat{G}^2 \sum_{j=1}^{N-1} u^2(j) \tag{2-25}$$

若所分析的信号$s(n)$符合公式(2-11)所定义的模型，那么输入信号$u(n)$可以认为是一个单位方差的白噪声序列。如果只考虑$s(n)$被一短时窗截得的部分，那么输入信号也可以是一个单位脉冲序列$\delta(n)$。在这种情况下，可以得出：

$$\hat{G} = \hat{E}_n^{1/2} \tag{2-26}$$

事实上，语音信号可以近似认为由清音和浊音组成的信号，对于浊音，激励$e(n)$是以基音周期重复的单位冲激；对于清音，$e(n)$接近白噪声，所以上述模型的假定能获得较好的效果[42]。

（2）求解线性预测方程组的程序实现

由前面介绍的线性预测原理可知在建立说话人识别模型的同时确定了线性预测系数为变量的线性方程组，即公式(2-22)，重新将其定义如下：

$$\sum_{k=1}^{p} a_k \times \phi_n(i,k) = \phi_n(i,0) \qquad i = 1,2,\cdots,p \tag{2-27}$$

其中$\phi_n(i,k)$只给出了以下形式：

$$\phi_n(i,k) = \sum_j s_n(j-i) \times s(j-k) \tag{2-28}$$

上式计算$\phi_n(i,k)$中j的求和范围没有给定。通常$\phi_n(i,k)$可以定义为自相关函数，方程组(2-24)有多种解法，以下给出*Durbin*递推算法[43]，该递推程序如下：

（1）给定预测器阶数 p；

（2）计算短时自相关函数 $R(l), l=0,1,\cdots,p$；

（3）计算 $K^{(1)}=-R(1)/R(0)$；

（4）计算 $a_1^{(1)}=K^{(1)}$；

（5）$\sum_{\beta}^{(1)}=\left[1-\left\{K^{(1)}\right\}^2\right]R(0)$；

（6）令 $m=2$；

（7）$K^{(m)}=-\left[-R(m)+\sum_{i=1}^{m-1}a_i^{m-1}R(|i-m|)\right]/\sum_{\beta}^{(m-1)}; a_m^m=K^{(m)}$；

（8）$a_i^m=a_i^{m-1}+K^{(m)}a_{m-i}^{m-1}, i=1,2,\cdots,(m-1)$；

（9）$m<p$？若回答是，则令 $m=m+1$，转入（7）继续执行；若回答否，则停止运行并输出 $a_1^p, a_2^p, \cdots, a_p^p$ 作为最后计算结果。

每帧提取 12 个 LPC 系数，对每个说话人的短时自相关函数进行归一化处理，变成小于 1 的数，这样可以便于进行数据处理和存储。其具体流程如下图所示：

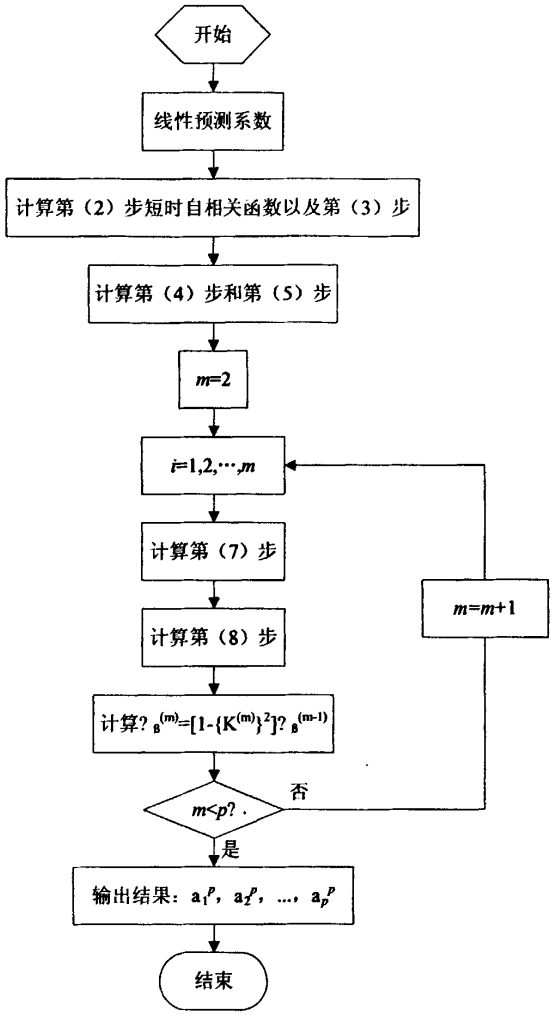

图 2-2 LPC 系数德宾（Durbin）递推法流程图

2.3.2 LPC 倒谱参数

线性预测[39] (Linear Prediction, LP)分析是最有效的语音分析技术之一，在语音编码、语音合成、语音识别、说话人识别等语音处理领域得到了广泛应用。线性预测分析的基本思想是：语音信号样点之间存在相关性，可以用过去的若干个样点或它们的线性组合预测现在或将来的样点值。可以通过使实际语音抽样值和线性预测抽样值之间的均方误差最小，得到一组唯一的线性预测系数

（LPC 系数）[17]。线性预测分析不仅能够提供语音信号的预测波形，而且能够提供一个好的声道模型。语音线性预测系数作为语音信号的一种特征参数，已被广泛应用于语音处理的各个领域。在 LPC 系数、LPC 反射系数、LPC 自相关函数、LPC 面积函数和 LPC 倒谱系数等多种 LPC 语音特征量中，倒谱系数对说话人识别效果最好。在对语音的浊音帧和清音帧特征参数的分析中发现，清音帧类似噪音，能量较低，易受背景噪音影响，而浊音帧的能量和规律性都较强，特征参数包含更多的说话人个体信息，是说话人识别研究的主要对象。

线性预测系数是线性预测的基本参数，可以将这些参数进行变换得到语音信号的其他参数，由线性预测系数得到线性预测倒谱系数的过程如下。

设通过线性预测分析得到的声道模型的系统函数为：

$$H(z)=\frac{1}{A(z)}=\frac{1}{1+\sum_{i=1}^{p}a_i z^{-i}} \tag{2-29}$$

其冲激响应为 $h(n)$，此处计算其倒谱 $h'(n)$。根据倒谱的定义，

$$\ln H(z)=H'(z)=\sum_{n=1}^{+\infty}h'(n)z^{-n} \tag{2-30}$$

将式(2－29)代入式(2－30)，并将其两边对 z^{-1} 求导数，即有：

$$\frac{\partial}{\partial z^{-1}}\ln\left[\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}}\right]=\frac{\partial}{\partial z^{-1}}\sum_{n=1}^{+\infty}h'(n)z^{-n} \tag{2-31}$$

即：

$$\sum_{n=1}^{+\infty}nh'(n)z^{-n+1}=\frac{\sum_{k=1}^{p}ka_k z^{-k+1}}{1-\sum_{k=1}^{p}a_k z^{-k}} \tag{2-32}$$

因而有：

$$\left(1-\sum_{k=1}^{p}a_k z^{-k}\right)\sum_{n=1}^{+\infty}nh'(n)z^{-n+1}=\sum_{k=1}^{p}ka_k z^{-k+1} \tag{2-33}$$

令其左右两边的常数项和 z^{-1} 各次幂的系数分别相等，即得到 $h'(n)$ 和 a_k 之

间的递推关系：

$$\begin{cases} h'(0)=0 \\ h'(1)=a_1 \\ h'(n)=a_n+\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)a_k h'(n-h) & 1\le n\le p \\ h'(n)=\sum_{k=1}^{p}\left(1-\frac{k}{n}\right)a_k h'(n-k) & n>p \end{cases} \tag{2-34}$$

由于 LPC 阶数 P 一般取 14，要小于一帧语音采样点数 N_s，因此 LPCC 只代表 $h'(n)(n=1,2,\cdots,N_s)$ 的前 P 个值。若倒谱分析阶数大于 P 时，由式(2-34)的第四部分即可求出。实验发现倒谱分析阶数取 16 能较好地表征语音的特征参数。

LPCC 反映的是说话人声道特征。这个倒谱是从一帧短时语音段中获取的，是语音在某一时刻某一帧的倒谱。它反映了语音信号倒谱的静态信息，故称为静态倒谱。由于语音信号的缓变特性，任意时刻的某一帧倒谱将有所不同，即静态倒谱将随时间作缓慢变化，这个变化的轨迹即倒谱的动态信息。短时谱随时间的变化表示为：

$$\sum_{n=-\infty}^{+\infty}\frac{dh'(n)(t)}{dt}e^{-jwm} \tag{2-35}$$

上式只能用有限差分近似，利用在有限长窗函数内的多项式来拟合倒谱系数的轨迹。一阶正交多项式系数，即时间上的广义谱斜率可表示为 $\Delta h'(n)(t)$。

$$\frac{dh'(n)(t)}{dt}\approx\Delta h'(n)(t)=\sum_{k=-K}^{K}kW_K h'(n)(t+k)/\sum_{k=-K}^{K}W_K k^2 \tag{2-36}$$

其中 W_K 是长为 $2K+1$ 的窗，$\Delta h'(n)(t)$ 称为动态倒谱。

由于选用的两种倒谱一个反映了静态信息，另一个反映了动态信息，两者互相补偿，充分表征了说话人声道模型。

语音的基音频率是声带振动的基本频率，它反映了声带激励源的特点。基音容易被模仿，不宜单独使用，但它可以与倒谱参数相结合。由于倒谱参数和基音参数分别描述了说话人声道、声带特征，从而可以充分反映说话人特征。

LPCC 的各种变形，例如差分倒谱、倒谱加权、自适应分量加权倒谱、倒谱均值减、ARMA 模型的零极点倒谱、RASTA 倒谱等也已成功地应用在噪声语音特征提取中。

2.3.3 MEL 频率倒谱参数

LPC 模型是基于语音发音机理的，描述的是声道特性，LPCC 系数也是基于合成的参数，这种参数没有充分利用人耳的听觉特性。在语音识别中，常用的语音特征是基于 Mel 频率的倒谱系数(Mel Frequeney Cepstrum Coefficient, MFCC)。由于 MFCC 参数是将人耳的听觉感知特性和语音的产生机制相结合，因此目前大多数语音识别系统中广泛使用这种特征。

人耳具有一些特殊的功能，这些功能使得人耳在嘈杂的环境中，以及各种变异情况下仍能正常地分辨出各种语音，其中耳蜗起了很关键的作用。耳蜗实质上相当于一个滤波器组，耳蜗的滤波作用是在对数频率尺度上进行的，在 1000Hz 以下为线性尺度，而 1000Hz 以上为对数尺度，这就使得人耳对低频信号比对高频信号更敏感。根据这一原则，研究者根据心理学实验得到了类似于耳蜗作用的一组滤波器组，这就是 Mel 频率滤波器组。

Mel 频率倒谱系数是将信号的频谱，首先在频域将频率轴变换为 Mel 频率刻度，再变换到倒谱域得到的倒谱系数。

Mel，是音高的单位，音高是一种主观心理量，是人类听觉系统对声音频率的感觉，Mel 频率刻度与频率的关系是：

$$Mel = \ln\left(1+\frac{f}{700}\right)\times\frac{1000}{\ln\left(1+\frac{1000}{700}\right)} \tag{2-37}$$

在实际应用中，MFCC 的计算过程如下：

（1）将信号进行短时傅里叶变换得到其频谱。

（2）求它的频谱幅度的平方，即能量谱，并用一组三角形滤波器在频域对能量谱进行带通滤波。这组带通滤波器的中心频率是按 Mel 频率刻度均匀排列的(间隔 150Mel，带宽 300Mel)，每个滤波器的三角形的两个底点的频率分别等于相邻的两个滤波器的中心频率，即每两个相邻的滤波器的过渡带相互搭接，且滤波器数为 M，滤波后得到的输出为：

$$X(k), k = 1,2,\cdots,M; \tag{2-38}$$

（3）将滤波器组的输出取对数，然后对它作 $2M$ 点逆离散傅里叶变换即得到 MFCC。由于对称性，此变换式可简化为：

$$C_n = \sum_{k=1}^{M} \log X(k) \cos\left[\pi(k-0.5)n/M\right], n = 1,2,K,L \qquad (2\text{-}39)$$

通常不用 0 阶倒谱系数，因为它是反映频谱能量的。

2.4 模式匹配方法

语音识别过程是根据模式匹配原则，计算未知语音模式与语音模板库中的每一个模板的距离测度，从而得最最佳的匹配模式[18,19]。要建立一个性能好的语音识别系统仅有好的语音特征是不够的，还要有好的语音识别的模型以及测度估计算法。

2.4.1 矢量量化与失真测度

矢量量化(Vector Quantization, VQ)技术是七十年代后期发展起来的一种数据压缩和编码技术，广泛应用于语音编码、语音合成、语音识别和说话人识别等领域。矢量量化在语音信号处理中占有十分重要的地位。在语音识别方面，矢量量化技术和动态时间规整(DTW)、隐马尔可夫模型(HMM)、人工神经网络(ANN)等方法的结合，提出了各种有效的识别方法。

如下图 2-3 所示，矢量量化技术在语音识别中的应用时，一般是先用矢量量化的码书作为语音识别的参考模板，即为系统中的每一个语音建立一个码书作为该语音的参考模板。识别时对于任意输入的语音特征矢量序列 $X_1, X_2, \cdots, X_N$，计算该序列对每个码书的总平均失真量化误差，即语音每一帧特征矢量与码书的失真之和除以该语音长度(帧数)。总平均失真误差最小的码书所对应的语音即为识别结果。

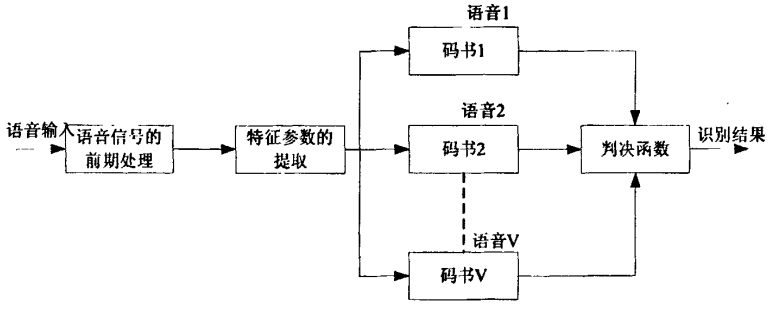

图 2-3 矢量量化在语音识别中的应用

利用矢量量化技术时，主要有以下两个问题[11]：

（1）设计一个好的码书。关键是如何划 W 个区域边界。这需要用大量的输入信号矢量，经过统计实验才能确定，这个过程称为“训练”或“学习”，其任务是建立码书。应用聚类算法，按照一定的失真度准则，对训练数据进行分类，从而把训练数据在多维空间划分成一个个以型心(码字)为中心的包腔，常用 LBG 算法来实现。

为了建立一个好的码书，首先要求建立码书的训练数据不仅数据量要充分大，而且要有代表性；其次，要选择一个好的失真测度准则及码本优化方法。

（2）未知矢量的量化。对未知模式矢量，按照选定的失真测度准则，把未知矢量量化为失真测度最小的区域边界的中心矢量值(码字矢量)，并获得该码字的序列号(码字在码书中的地址或标号)。被量化矢量与其对应的矢量存在一定的失真测度值，它描述了当输入矢量用码书中对应的码矢来表征时所付出的代价。

失真测度(距离测度)是将输入矢量 X_i，用码书重构矢量 Y_j 来表征时所产生的误差或失真的度量方法，它可以描述两个或多个模型矢量间的相似程度。失真测度的选择好坏将直接影响到聚类效果和量化精度，进而影响到语音信号矢量量化处理系统的性能。

设两个 M 维语音特征矢量 X 和 Y 进行比较，要使其距离测度 $d(X,Y)$ 在语音信号处理中有效，必须具备下列条件：

（1）对称性 $d(X,Y)=d(Y,X)$。

（2）正值性 $d(X,Y)\geq 0$，当且仅当 $X=Y$ 时等号成立。

（3）三角不等式 $d(X,Y)\leq d(X,Z)+d(Z,Y)$。

（4）与语音质量的主观评价相一致。

（5）易于计算。

在语音信号处理采用的矢量量化中，最常用的失真测度是欧氏距离测度、加权欧氏距离测度、似然比失真测度和识别失真测度。本课题采用的是欧氏距离测度，下面将具体介绍下欧氏距离测度。

设未知模式的 M 维特征矢量为 X，与码书中某个 M 维码矢 Y 进行比较，x_i 和 y_j 分别表示 X 和 Y 的同一维分量($1\leq i\leq M$)，则几种常用的欧氏距离测度如

下：

（1） 均方误差欧氏趾离。其定义为：

$$d_2(X,Y)=\frac{1}{M}\sum_{i=1}^{M}\left(x_i-y_j\right)^2=\frac{(X-Y)^T(X-Y)}{M} \tag{2-40}$$

这里的$d_2(X,Y)$下标 2 表示平方误差。

（2）r方平均误差。其定义为：

$$d_r(X,Y)=\frac{1}{M}\sum_{i=1}^{M}(x_i-y_j)^r \tag{2-41}$$

（3）r平均误差。其定义为：

$$d_r'(X,Y)=\left[\frac{1}{M}\sum_{i=1}^{M}|(x_i-y_j)|^r\right]^{\frac{1}{r}} \tag{2-42}$$

（4）加权欧氏距离测度。如下定义加权欧氏距离测度：

$$d(Z,Y)=\frac{1}{M}\sum_{i=1}^{M}w(i)(x_i-y_j)^2 \tag{2-43}$$

其中，$w(i)$称为加权系数。

2.4.2 动态时间规整技术

动态时间规整[8]（DTW）采用动态规划技术(Dynamic Programming, DP)将一个复杂的全局最优化问题化为许多局部最优化问题一步一步地进行决策。如设参考模板特征矢量序列为$A=\{a_1,a_2,\cdots,a_I\}$，输入语音特征矢量序列为$B=\{b_1,b_2,\cdots,b_J\}$。

上面$I\neq J$，那么 DTW 算法就是要寻找一个最佳的时间规正函数，使被测语音模板的时间轴i，非线性地映射到参考模板的时间轴j，使总的累计失真量最小，设时间规正函数为：

$$C=\{c(1),c(2),\cdots,c(N)\} \tag{2-44}$$

其中N为路径长度，$c(n)=\left(i(n),j(n)\right)$表示第$n$个匹配点对，它是由参考模板的第$i(n)$个特征矢量与被测模板的第$j(n)$个特征矢量构成的匹配点对。二者之间的距离(或失真值)$d\left(a_{i(n)},b_{j(n)}\right)$称为局部匹配距离。DWT 算法就是通过局部优化的方法实现加权距离总和最小，即：

$$D=\min\frac{\sum_{n=1}^{N}\left[d\left(a_{i(n)},b_{j(n)}\right)\times W_n\right]}{\sum_{n=1}^{N}W_n} \tag{2-45}$$

其中加权函数的选取可考虑两个因素：（1）根据第n对匹配点前一步局部路径的走向来选取，去除 45 度方向的局部路径，以便适应$I\neq J$的情况；（2）考虑语音各部分给不同权值以加强某些区别特征。在公式(2-45)所表达的优化过程中，对时间规正函数 C 作某些限定，以保证匹配路径不违背语音信号部分特征的时间顺序。一般要求规正函数满足如下约束：

（1）必须是单调的：$i(n)\geq i(n-1), j(n)\geq j(n-1)$；

（2）起点和终点约束：一般要求$i(1)=j(1)=1; i(N)=I, j(N)=J$；

（3）连续性：一般规定不允许跳过任何一点，即：$i(n)-i(n-1)\leq 1$和$j(n)-j(n-1)\leq 1$；

（4）最大规正量不超过某一极限，最简单的情形为$|i(n)-j(n)|\leq M$，其中M为窗宽。

通常还对规正函数所处的区域作某些规走，如位于平行四边形内，为了实现以上约束条件，需要设计局部路径的约束，它用于限制第n步为$(i(n), j(n))$时前几步存在几种可能的局部路径[8,20]。

图 2—4 给出了三种典型的局部路径约束，(a)，(b)，(c)分别示出了路径受前面一步、二步和三步约束的情况。

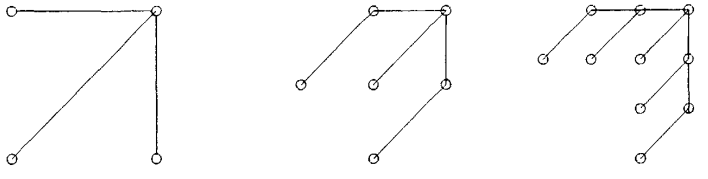

图 2-4 三种典型的局部路径约束

下面再定义一种最小累计失真函数$g(i,j)$，表示匹配点对(i,j)为止，前面所有可能的路径中最佳路径的累计匹配距离$g(i,j)$存在如下递推关系：

$$g(i,j)=\min_{(i',j')\to(i,j)}\left\{g(i',j')+d(a_i,b_j)w_n\right\} \tag{2-46}$$

其中(i',j')表示局部路径$(i',j')\to(i,j)$的起点，权w_n的取值是与局部路径有关的。

下面基于上述定义、约束和规则，并以图 2-4 的局部路径约束和平行四边形区域约束为例介绍 DTW 算法的具体步骤：

（1）初始化：

令$i(1)=j(1)=1, g(1,1)=2d(a_1,b_1)$

$$g(i,j)=\begin{cases}0 & when \quad (i,j)\in \mathrm{Re}g\\ huge & when \quad (i,j)\notin \mathrm{Re}g\end{cases} \tag{2-47}$$

其中约束区域Reg可假定它是这样一个平行四边形，它有两个顶点位于(1,1)和(I,J)，相邻两条边的斜率分别为 2 和 1/2。

（2）递推求累积距离：

$$\begin{aligned}g(i,j)=\min\{&g(i-1,j)+d(a_{i-1},b_j)\times W_n(1);\\ &g(i-1,j-1)+d(a_i,b_j)\times W_n(2);\\ &g(i,j-1)+d(a_i,b_j)\times W_n(3)\}\end{aligned} \tag{2-48}$$

对于图 2-4 所示的局部路径，一般取距离加权值为$W_n(1)=W_n(3)=1$，$W_n(2)=2$，规正函数的点数不是固定不变的，随I和J的值而变，这可以用W_n作为分母来补偿，如公式公式(2-49)所示。不难证明，当距离加权函数取得合适时，有：

$$\sum_n W_n=I+J=\text{常数} \tag{2-49}$$

于是求得最终的匹配加权距离：

$$D=g(I,J)/(I+J) \tag{2-50}$$

（3）回溯求出所有的匹配点对：根据上一步每步的最佳局部路径，由匹配点对(I,J)向前回溯一直到(1,1)。这个回溯过程对于求平均模板或聚类中心来讲是必不可少的，但在识别过程往往不必进行[21]。

第 3 章　基于蚁群算法的动态时间规划算法设计

3.1　基本蚁群算法

3.1.1 蚁群算法简介

从 20 世纪 50 年代中期开始，仿生学日益得到人们的重视。受仿生学中生物进化机理的启发，人们提出了一系列新的算法，解决了许多比较复杂的优化问题。遗传算法、人工免疫算法、神经网络等算法相继出现，并得到了发展，逐渐成为比较成熟的算法。

在二十世纪九十年代初期，意大利 Dorigo M、Maniezzo V、Colorni A 等人从蚂蚁觅食的自然现象中受到启发，经过大量的观察和实验发现，蚂蚁在觅食过程中留下了一种外激素，又叫信息激素。它是蚂蚁分泌的一种化学物质，蚂蚁在寻找食物的时候会在经过的路上留下这种物质，以便在回巢时不至于迷路，而且方便找到回巢的最好路径。由此，Dorigo M 等人首先提出了一种新的启发式优化算法，叫蚁群算法(ACA)[22]。蚁群算法是最新发展的一种模拟昆虫王国中蚂蚁群体智能行为的仿生优化算法，它具有较强的鲁棒性、优良的分布式计算机制、易于与其他方法相结合等优点。

蚁群算法的主要特点是：正反馈、分布式计算。正反馈过程使得该方法能很快发现较好解，分布式易于并行实现，与启发式算法相结合，使得该方法易于发现较好解。

蚁群算法的出现为解决 NP 难度问题提供了一条新的途径。随后该方法解决了一系列组合优化问题，如旅行商问题（TSP）[23,24,25]、二次分配问题（QAP）[26,27,28]、车间调度（JSP）[29]、图着色（GCP）[30,31]等。虽然对蚁群算法研究的时间不长，但是初步研究已显示出蚁群算法在求解复杂优化问题（特别是离散优化问题）方面具有一定优势，表明它是一种很有发展前景的方法。

3.1.2 蚁群算法的产生

仿生学家经过长期研究发现在自然界真实的蚁群觅食过程中，蚂蚁虽然没有视觉，但可以在路径上释放一种特殊的分泌物——信息素（pheromone）来进

行个体之间的信息交换寻找路径。开始时蚂蚁会随机挑选一条路径前进。蚂蚁走的路径越长，释放的信息量越小。在寻找路径过程中，每只蚂蚁倾向于选择信息素浓度较大的路径。当一些路径上通过的蚂蚁越来越多时，该路径上的信息素浓度就越大，后来的蚂蚁选择该路径的可能性就越大，从而进一步增加了该路径上的信息素浓度，这种选择过程称为蚂蚁的自催化行为（auto-catalytic behavior），其原理可以看成一种正反馈机制。从而最优路径上的信息素浓度会不断增强，其他路径上的信息素会随着时间的流逝而逐渐减弱，最终整个蚁群会找到最优路径。同时蚁群还能适应环境改变找到当前的最佳路径，如图 3-1 所示。

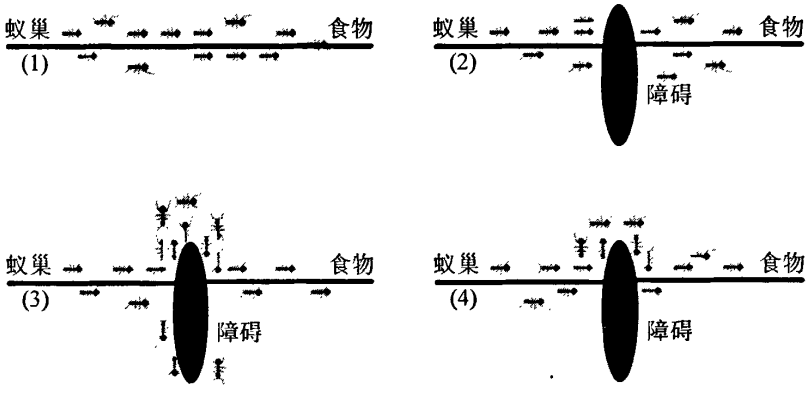

图 3-1 蚁群觅食过程

在图 3-1 中，形象的显示了蚁群觅食的过程：

（1）蚁群在蚁巢和食物之间建立通路；

（2）当在通道上出现一障碍物，蚁群会等概率的选择沿着障碍物向左或向右移动；

（3）蚁群会在较短路径上留下较多信息素以指导后面的蚁群移动；

（4）直至所有蚂蚁都选择同一较短路径。

3.2 基本蚁群算法模型的建立

3.2.1 蚁群算法的实现

蚁群算法[45]最成功的就是运用在旅行商[44]（traveling salesman problem，TSP）问题上[23]。TSP 具有广泛的代表意义和应用前景，许多问题均可抽象为

TSP 的求解。

TSP 就是指给定 n 个城市和两两城市之间的距离，要求确定一条经过各城市当且仅当一次的最短路线。设 m 是蚁群中蚂蚁的数量，用 d_{ij} 表示城市 i 和城市 j 之间的距离，$\tau_{ij}(t)$ 表示 t 时刻残留在城市 i、j 连线上的信息量。初始时刻 $t=0$ 时，将 m 只蚂蚁随机放置到 n 个城市中的 m 个城市上，各条路径上的信息素量相等，设 $\tau_{ij}(0)=C$（C 为常数）。蚂蚁 $k(k=1,2,3,\cdots,m)$ 在运动过程中根据各条路径上的信息素量决定转移方向。蚂蚁系统所使用的状态转移规则被称为随机比例规则，它给出了位与城市 i 的蚂蚁 k 选择移动到城市 j 的概率。在 t 时刻，蚂蚁 k 在城市 i 选择城市 j 的转移概率 $P_{ij}^{k}(t)$ 为：

$$P_{ij}^{k}(t)=\begin{cases}\dfrac{\left[\tau_{ij}(t)\right]^{\alpha}\times\left[\eta_{ij}(t)\right]^{\beta}}{\sum\limits_{allowed_k}\left[\tau_{ij}(t)\right]^{\alpha}\times\left[\eta_{ij}(t)\right]^{\beta}} & if\ \ j\in allowed_k \\ 0 & otherwise\end{cases} \tag{3-1}$$

其中，$allowed_k=\{0,1,\cdots,n-1\}$ 表示蚂蚁 k 下一步允许选择的城市。η_{ij} 为启发函数，其表达式如下：

$$\eta_{ij}=1/d_{ij} \tag{3-2}$$

d_{ij} 是两城市 i 和 j 之间的距离。α 和 β 为两个参数，分别反映了蚂蚁在运动过程中所积累的信息和启发信息在蚂蚁选择路径中的相对重要性。为了使蚂蚁经过 n 个不同的城市，每只蚂蚁都设计了一个数据结构，称为禁忌表（tabu list）。禁忌表记录了在 t 时刻蚂蚁已经走过的城市，不允许该蚂蚁在本次循环中再经过这些城市。当一次循环结束后，禁忌表被清空，该蚂蚁又可以自由地进行选择。

蚂蚁完成一次循环，各路径上的信息素量根据下式进行调整：

$$\tau_{ij}(t+n)=(1-\rho)\times\tau_{ij}(t)+\Delta\tau_{ij}(t) \tag{3-3}$$

$$\Delta\tau_{ij}(t)=\sum_{k=1}^{m}\Delta\tau_{ij}^{k}(t) \tag{3-4}$$

其中 $\rho<1$ 表示蚂蚁在 $t+n$ 时刻留在路径 (i,j) 上信息素量；$\Delta\tau_{ij}(t)$ 表示本次循环中路径 (i,j) 的信息素量的增量；$(1-\rho)$ 为信息素轨迹的残留因子，通常设置参数 $\rho<1$ 来避免路径上信息素量的无限累加。

算法实现的流程图如图 3-2 所示：

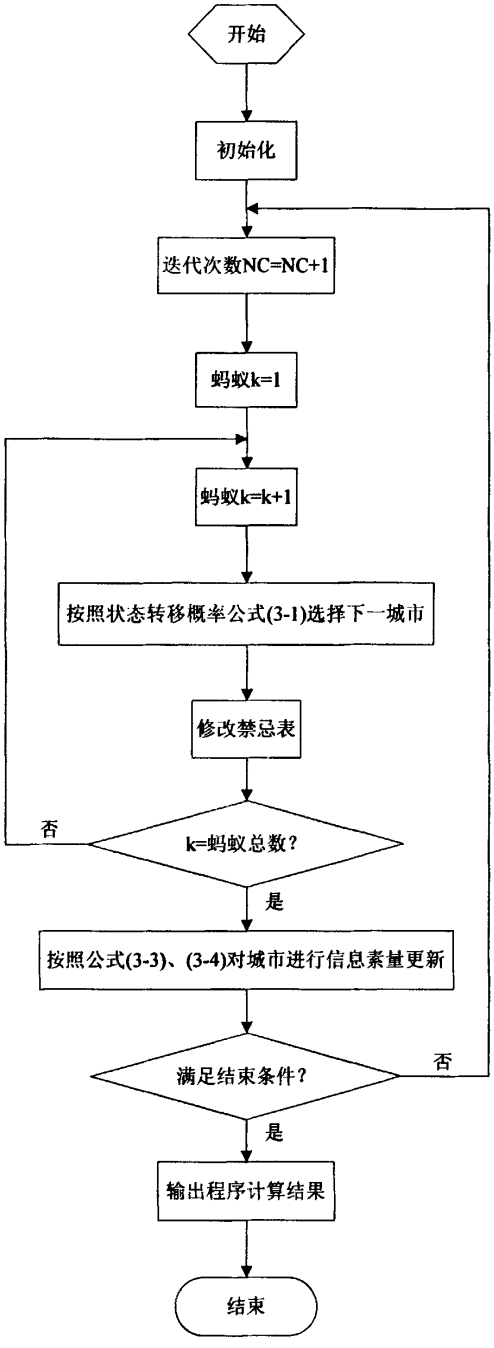

图 3-2 蚁群算法的程序结构流程图

3.2.2 蚁群算法的数学模型

根据信息素更新策略的不同，Dorigo M 提出了三种不同的基本蚁群算法模型，分别称之为蚁量系统（ant-quantity system）模型、蚁密系统（ant-density system）模型、蚁周系统（ant-cycle system）模型[32]。三种模型的差别仅在于的表达式不同。

在蚁量系统模型中，

$$\Delta\tau_{ij}^{k}(t)=\begin{cases}\dfrac{Q}{d_{ij}}, & \text{如果第}k\text{只蚂蚁在本次循环中经过路径}(i,j)\\ 0, & \text{否则}\end{cases} \tag{3-5}$$

式中，Q为信息素强度。从上式可见，蚁量系统中，一只蚂蚁在路径(i,j)上释放的信息素量为Q/d_{ij}，因此较短路径对蚂蚁更有吸引力。

在蚁密系统模型中，

$$\Delta\tau_{ij}^{k}(t)=\begin{cases}Q, & \text{如果第}k\text{只蚂蚁在本次循环中经过路径}(i,j)\\ 0, & \text{否则}\end{cases} \tag{3-6}$$

可见一只蚂蚁从i向j移动的过程中路径(i,j)上信息轨迹强度与d_{ij}无关。

在蚁周系统模型中，

$$\Delta\tau_{ij}^{k}(t)=\begin{cases}\dfrac{Q}{L_k}, & \text{如果第}k\text{只蚂蚁在本次循环中经过路径}(i,j)\\ 0, & \text{否则}\end{cases} \tag{3-7}$$

其中，L_k是第k只蚂蚁在本次循环中所走的路径长度。与蚁量和蚁密系统不同的是，蚁周系统是在蚂蚁已建立了完整的轨迹后再释放信息素，利用的是整体信息；而前面两者则是在建立方案的同时释放信息素，利用的是局部信息。在求解 TSP 问题时，蚁周算法模型明显优于其他两种算法模型，因此就常采用式(3-7)作为蚁群算法的基本模型。

3.3 蚁群算法复杂度的分析

通常我们用复杂度来表示算法执行效率的高低。将基本蚁群算法的复杂度表示为问题规模n（TSP 中的城市数目）的函数。通过对算法流程结构的分析，如果m只蚂蚁要遍历n个城市，经过N_c次循环，可逐步分析出其时间复杂度，

如表 3-1 所示。整个计算过程的时间复杂度为[24]：

$$T(n)=O(N_c\times n^2\times m) \tag{3-8}$$

而空间复杂度为：

$$S(n)=O(n^2)+O(n\times m) \tag{3-9}$$

但一般情况下 $m<<n$，因此整个算法的空间复杂度为 $O(n^2)$。可见，数据存储上基本蚁群算法的空间复杂度是简单的，易于程序编制。

表 3-1 基本蚁群算法的时间复杂度分析

步骤	内容	时间复杂度
1	初始化参数	$O(n^2+m)$
2	设置蚂蚁禁忌表	$O(m)$
3	每只蚂蚁单独构造解	$O(n^2m)$
4	轨迹更新量的计算	$O(n^2m)$
5	轨迹的信息素量的更新	$O(n^2)$
6	判断是否达到算法终止条件，否则转到第 2 步	$O(nm)$
7	输出计算结果	$O(1)$

3.4 蚁群算法的参数优化问题

蚁群算法中参数 α、β、ρ、Q 的取值直接影响到算法的全局收敛性和求解效率。蚁群算法的参数最优组合问题是一个极其复杂的优化问题，目前尚没有完善的理论依据。大多数论文都没有进行讨论，通常都是按照经验选取一组值，这些经验都是来源于试验的结果，纯粹从理论上分析的成果目前还很少。有一些论文对蚁群算法收敛性进行了分析[24,33,34]，但是目前这些分析的结果都还很难用于指导参数选择。在较早的一些蚁群算法论文中，就对各种参数的取值进行过试验验证，但是要想通过人工的方式来取各种不同的参数组合来验证是很困难的，因此目前都没有对参数组合作过多讨论。有许多学者研究过遗传算法与蚁群算法的融合，这些研究中有许多就是用遗传算法去优化蚁群算法的各个参数[35]。但对于不同问题，可能参数最优取值会有所不同。因此，蚁群算法中参数组合的问题还有待做进一步的深入研究。

3.5 蚁群动态时间规划算法

3.5.1 问题描述

在语音识别中的模式匹配过程中，由于语音信号特征参数序列往往是不等长的，所以就必须有一个好的测度估计算法来对不等长的序列进行测度估计，从而得到最佳的识别结果。DTW(dynamic time warping)是较早的一种模式匹配和模型训练技术，它应用动态规划方法成功解决了语音信号特征参数序列比较时时长不等的难题，在孤立词语音识别中获得了良好的性能[36]。DWT 算法在本课题第二章的 2.4.2 小节有详细分析，这里就不再作具体介绍。我们知道，DWT 算法是通过局部优化的方法实现加权距离总和最小的，是一种局部最优算法，其每一步搜索都是根据局部优化的判断进行的，因此这个时间规整路径达不到全局最优，且由于局部搜索是递归进行的，对全局路径的标准化及规整子路径的加权都无法全局衡量，必然使得识别结果较粗糙。而蚁群算法作为一种新的用于解决复杂优化问题的全局搜索方法，已经成功应用于求解 TSP 问题、调度问题、指派问题等，显示出了蚁群算法在处理复杂优化问题[3,4]方面的优越性。蚁群算法具有分布式计算、信息正反馈和启发式搜索的特征，本质上是进化算法中的一种新型随机性优化算法。

本课题中，我们利用蚁群算法优化机制，结合传统的 DTW 算法，提出了一种新的基于蚁群算法的动态时间规划算法来搜索语音信号特征参数序列之间匹配的一条全局最优路径，进而以此衡量语音信号之间的相似度，从而期望得到最佳的识别结果。

3.5.2 算法原理

蚁群算法实际上是一类智能多主体系统，其自组织机制使得蚁群算法不需要对所求问题的每一方面都有详尽的认识。自组织本质上是蚁群算法机制在没有外界作用下使系统熵增加的动态过程，体现了从无序到有序的动态演化[37]，其逻辑结构如图 3-3 所示。

由图 3-3 可见，先将具体的组合优化问题表述成规范的格式，然后利用蚁群算法在“探索（exploration）”和“利用（exploitation）”之间根据信息素这一反馈载体确定决策点，同时按照相应的信息素更新规则对每只蚂蚁个体的信息

素进行增量构建，随后从整体角度规划出蚁群活动的行为方向，周而复始，即可求出组合优化问题的最优解。

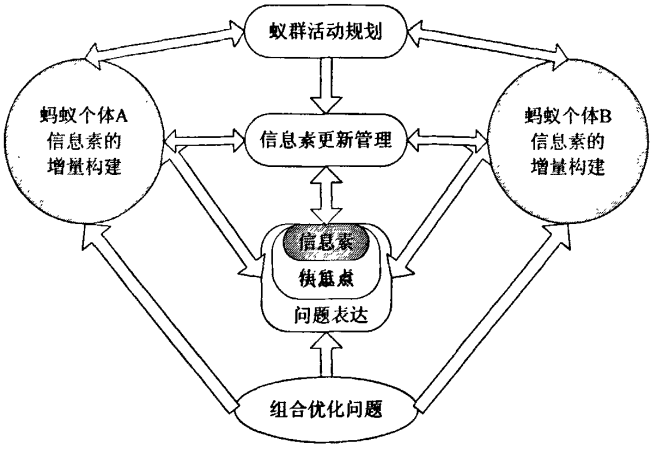

图 3-3 蚁群算法的逻辑结构

根据这个思路，在基于模式匹配的语音识别系统中，未知语音的模式要与已知参考语音的参考模式逐一进行比较，最佳匹配的参考模式作为识别结果输出。

参考模板可以表示如下：

$$R=\{R(1),R(2),\cdots,R(m),\cdots,R(M)\} \tag{3-10}$$

测试模板可以表示如下：

$$T=\{T(1),T(2),\cdots,T(n),\cdots,T(N)\} \tag{3-11}$$

R 与 T 之间的总体失真为 $D[R,T]$，失真越小，相似度越高。一般情况下 $M\neq N$。将 R、T 相应特征矢量映射为 $i-j$ 平面上的点，形成 $M\times N$ 个节点的网格。蚁群动态时间规划的目的是找到一条最优路径使得语音信号两模板 R 与 T 之间的匹配距离最小。

设 $C=\{c(1),c(2),\cdots,c(k),\cdots,c(K)\}$ 为待规划路径，K 为匹配路径的长度。$c(n)=(i(n),j(n))$ 为参考模板的第 $i(n)$ 个特征矢量与测试模板的第 $j(n)$ 个特征矢量构成的匹配点对。匹配起点为坐标点 $s=(1,1)$，终点为 $e=(M,N)$，(M,N) 匹配路径约束条件需满足以下几点：

（1）路径通过起点$(1,1)$和终点(M,N)。

（2）路径不允许跳过任何一点,即满足连续性。

（3）$m(k)\geq m(k-1),n(k)\geq n(k-1)$，即满足单调性。

（4）最大归整量不超过某一极限，不允许时间轴极度变化，即$|m(k)-n(k)|\leq r,r$称为窗宽。

根据以上 4 个约束条件，蚂蚁k在当前时刻位于点$m(i(n),j(n))$，则下一点所允许集合为：

$$allowed_k=\{(i(n),j(n+1)),(i(n),j(n)),(i(n+1),j(n+1))\} \tag{3-12}$$

设t时刻蚂蚁k从当前点m到下一点n的期望程度为$\eta_{mn}^{\alpha}(t)=\{1/[\eta_m(t)+d_n(t)]\}^{\beta}$，$n\in allowed_k$，信息量为$\tau_{mn}^{\alpha}(t)$，在满足约束条件(3-12)时，定义$t$时刻蚂蚁$k$从当前点$m(i(r),j(r))$到下一点$n$的期望程度为：

$$\eta_{mn}^{\alpha}(t)=\{1/[\eta_m(t)+d_n(t)]\}^{\beta},n\in allowed_k \tag{3-13}$$

其中，$\eta_m(t)$为语音信号在m点的局部累计匹配距离；$d_n(t)$为点n相对应语音矢量之间的匹配距离。

在满足约束条件式(3-12)时蚂蚁k从当前点m到下一点n的状态转移概率为：

$$P_{mn}^{k}(t)=\begin{cases}\dfrac{[\tau_{mn}(t)]^{\alpha}\times[\eta_{mn}]^{\beta}}{\sum_{r\in allowed_k}[\tau_{mn}(t)]^{\alpha}\times[\eta_{mn}]^{\beta}} & n\in allowed_k\\ 0 & otherwise\end{cases} \tag{3-14}$$

其中，$\tau_{mn}(t)$为蚂蚁k在t时刻搜索时，m、n之间路径的信息素强度；η_{mn}为由m到n的路径期望程度，η_{mn}越大，$P_{mn}^{k}(t)$越大。

蚁群动态时间规划算法中，每只蚂蚁从始点到终点根据状态转移概率随机进行路径搜索，同时对蚂蚁所走路径做严格约束，必须满足条件(3-12)且必须在一个上边斜率为 2 下边斜率为 1/2 的平行四边形内。当蚁群中所有蚂蚁完成一次循环后，更新路径上的信息素。然后，进行第二此循环，并再次路径上的信

息素，重写状态转移概率。经过若干次循环，最终根据路径上的信息素强度选择语音信号匹配最优路径。

3.5.3 信息素更新机制

针对语音识别系统，为了考虑匹配路径的全局信息，在信息素更新规则中引进代价函数 $D_{se}^{k}(t)$，定义为蚂蚁 k 从 s 到 e 规划路径 C 的全局平均匹配失真距离。

设语音信号匹配点对 $c(n)=(i(n),j(n))$，二者之间的距离(失真度)为 $d\left(R_{i(n)},T_{j(n)}\right)$，则 $D_{se}^{k}(t)$ 定义为：

$$D_{se}^{k}(t)=\left[\sum_{k=1}^{K}d\left(R_{i(n)},T_{j(n)}\right)\right]/K \tag{3-15}$$

$\Delta\tau_{mn}^{k}(t)$ 定义为：

$$\Delta\tau_{mn}^{k}(t)=\begin{cases}1/D_{se}^{k}(t) & \text{蚂蚁}k\text{从}s\text{到}e\text{路径}\\ 0 & \text{其它}\end{cases} \tag{3-16}$$

信息素更新为：

$$\tau_{mn}(t+1)=\rho\times\tau_{mn}(t)+\sum_{k=1}^{K}1/D_{se}^{k}(t) \tag{3-17}$$

从式(3-17)可知，$D_{se}^{k}(t)$ 越小，相应路径上的信息素增量就越大，从而引导蚁群选出语音信号匹配的最优路径。

由以上分析可知，蚁群动态时间规划算法利用了模板匹配过程中的局部信息 $\eta_{mn}(t)$，并根据全局代价函数 $D_{se}^{k}(t)$ 更新路径上的信息素值。因此，应用此算法得到的是全局最优解，且搜索过程中不必像 DTW 算法那样考虑权系数问题。

第 4 章　采用蚁群算法的语音识别系统

4.1　语音识别系统的软件实现

语音识别系统的工作流程图如图 4-1 所示，预加重参数取α =0.98，采用汉明窗进行信号的分帧，语音的端点检测采用短时平均能量与短时过零率相结合的方法，经过特征提取，在模式匹配中采用蚁群动态时间规划算法进行识别，最后得出识别结果。

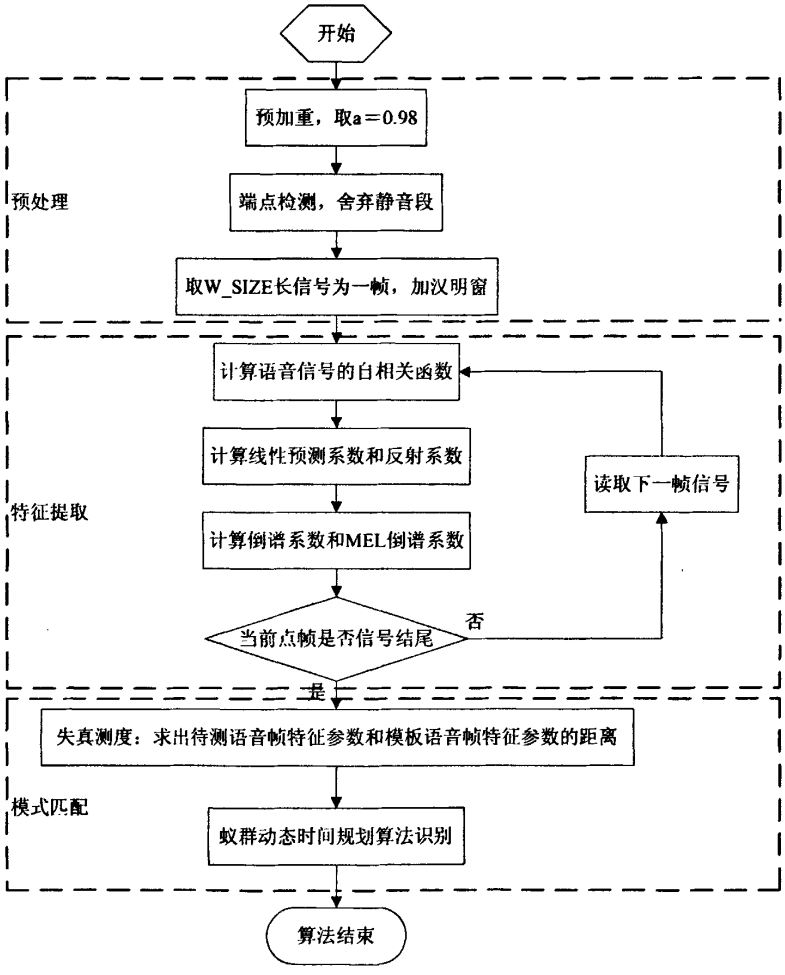

图 4-1 语音识别系统程序流程图

其中，语音信号的预处理的仿真、特征提取方法的仿真和模式匹配的仿真

将在下面几节介绍。

4.2 语音信号的预处理

语音信号获取是语音识别的第一步。作为语音的发出者，必须清晰准确的发音，以获得准确的语音样本，为以后的工作打好基础。语音样本的采集是应用 WINDOWS 自带的录音机附件来完成的。在采集过程中，将直接剔除那些明显被偶然因素干扰和因说话人本身造成的不规则样本。

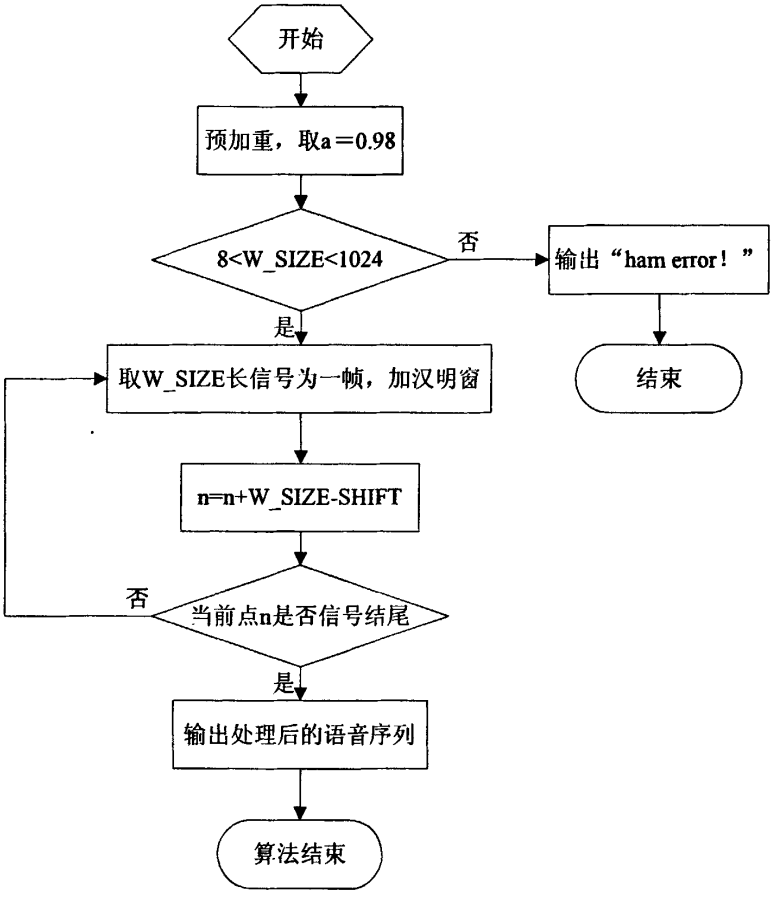

图 4-2 语音信号的预处理程序流程图

图 4-2 是本课题设计的语音识别系统信号的预处理程序流程图。本课题采用α=0.98 进行系统的预加重，采用汉明窗对信号进行分帧，另外设置了一个窗

长的范围 8<W_SIZE<1024，因为如果帧设置太短，不足于表现出语行特征，太长则违反了短时平稳的假设，当窗长超出这个范围，系统输出“ham error!”并结束程序。图 4-3 是男声发音“播放”在本课题设计的语音识别系统中经过预加重和汉明窗分帧前后的信号波形图。

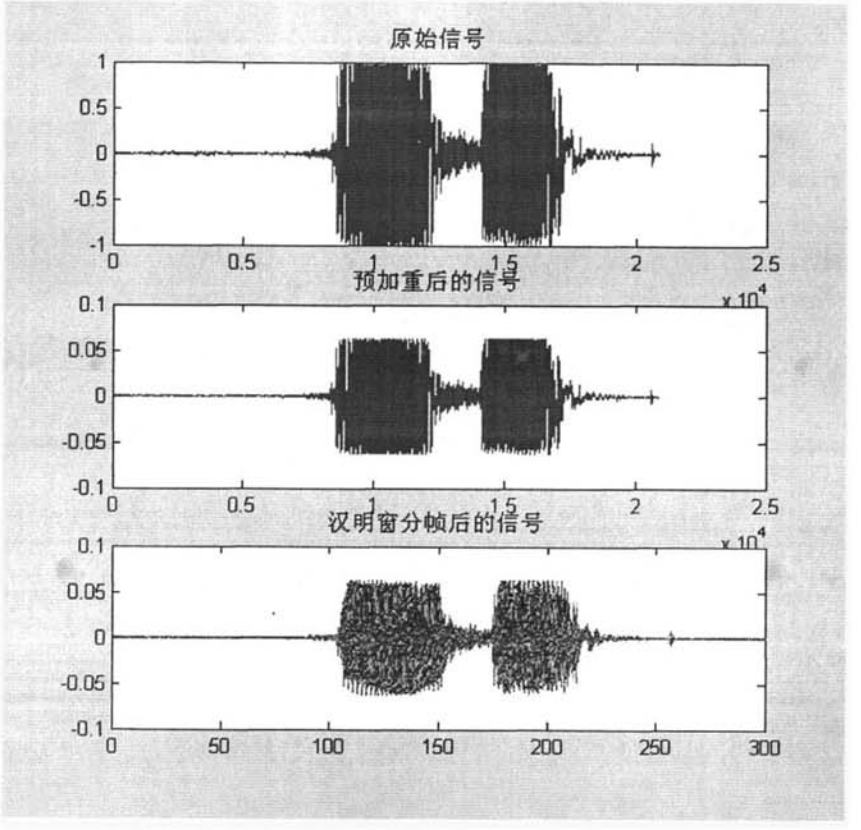

图 4-3 男声发音“播放”的原始信号与预加重、汉明窗分帧后的信号波形图

在端点检测部分，本课题采用短时能量和短时过零率相结合的方法，利用短时能量和短时过零率两个门限来确定语音信号的起点和终点，目的是从采集到的语音信号中分离出真正的语音信号作为系统处理的对象。

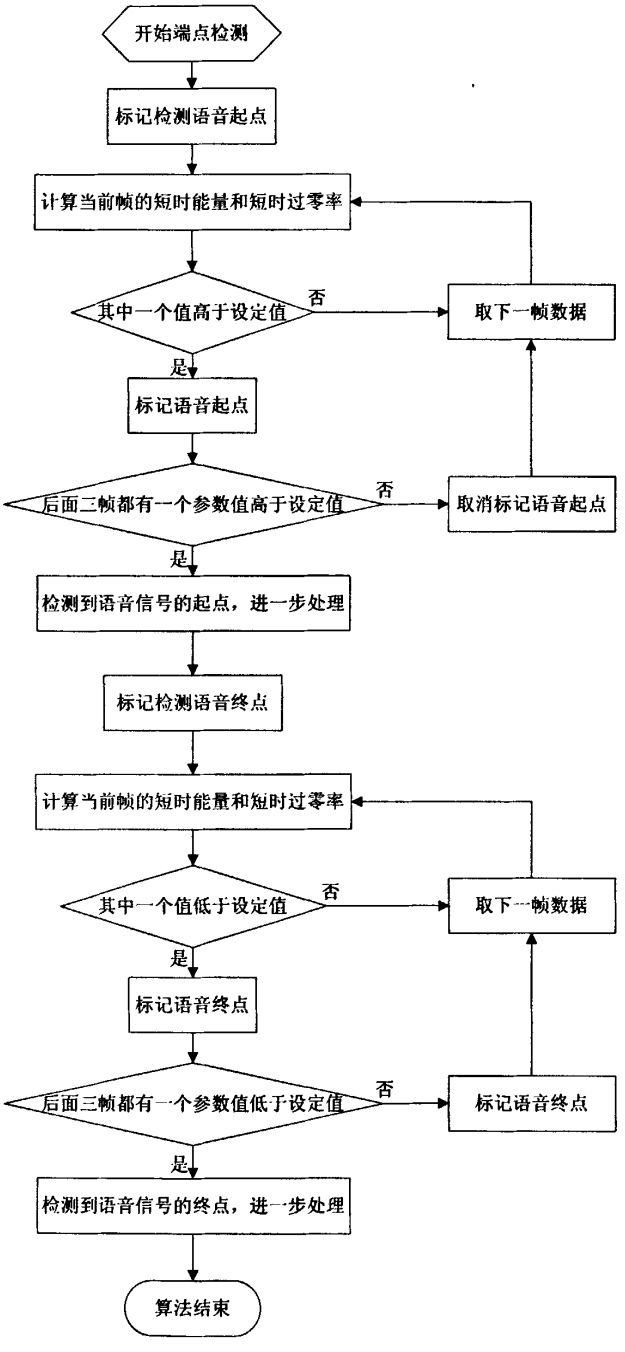

图 4-4 双门限端点检测程序流程图

端点检测程序流程图如图 4-4 所示，在语音信号端点检测前，先要求为短

时平均能量和短时过零率确定两个门限。在静音段，如果能量或过零率超越了低门限，就应该开始标记起始点，进入过渡段。在过渡段中，由于参数的数值比较小，不能确信是否进入语音段，只要两个参数的数值都回落到低门限以下，就将当前状态恢复到静音状态。而如果在过渡段中两个参数中的任一个超过了高门限，就可以确信进入语音段，就可以标记一段语音。

在检测到语音段后，标记开始检测语音终点，如果检测到短时能量或者短时过零率低于阈值，则标记为语音终点，进入过渡段，在过渡段中，由于参数的数值比较大，不能确信是否进入静音段，如果在过渡段中两个参数中的任一个超过了高门限，就可以确信还是语音段，继续标记语音，取下一帧再进行判断；只要两个参数的数值都回落到低门限以下，就将当前状态恢复到静音状态。识别效果如图 4-5 所示。

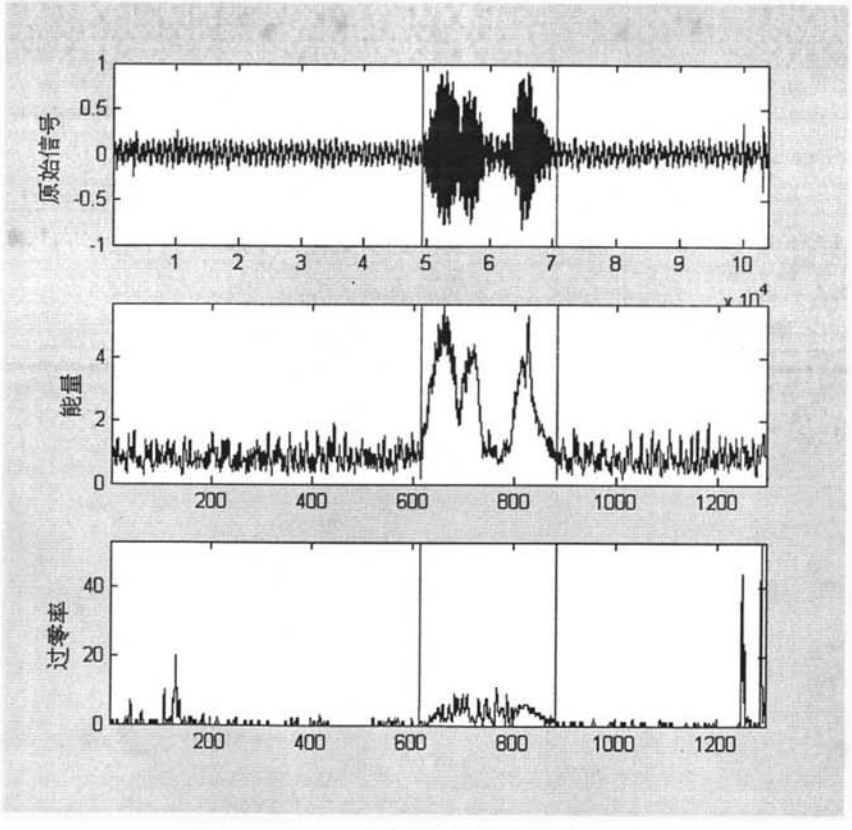

图 4-5 女声发音“你好”的端点检测示意图

4.3 语音信号的特征提取

特征矢量[11]的提取在语音识别中占有极其重要的地位，特征矢量提取得是否得当直接影响着语音识别率，因此必须给予足够的重视。特征矢量的提取是对原始的语音信号运用一定的数字信号处理技术进行适当的处理，从而得到一个矢量序列，这个矢量序列可以代表原始的语音信号所携带的信息，初步实现数据压缩。提取特征矢量的原则是：要尽可能保留那些对识别率有重要意义的特征信息，同时最大限度地摒弃那些对语音识别无用的冗余信息[38]。

基本的特征参数主要有：能量、幅度、过零率、频谱、倒谱和功率潜等，另外考虑到其他因素的影响，还有许多基于基本参数的参数，如从听觉出发，用来表达语音的特征有：MEL 频率倒谱系数(MFCC)、感知线性预测系数(PLP)等，这些参数相对于 LPC 或 FFT 等基本分析方法有许多优点[39]。本课题对语音信号采用了 12 阶 MEL 频率倒谱系数(MFCC)进行特征提取。

与普通实际频谱倒谱分析不同，MEL 频率倒谱参数的分析基于人耳的听觉特性[40]。因为，人耳听到的声音的高低与声音的频率并不成线性正比关系，Mel 频率尺度更符合人耳的听觉特性。所谓 Mel 频率尺度，它的值大体上对应于实际频率的对数分布关系，具体关系可用下式表示：

$$Mel(f) = 2595\log(1 + f/700) \tag{4-1}$$

实际频率 f 的单位是 Hz，临界频率带宽随着频率的变化而变化，并与 Mel 频率的增长一致。在 1000Hz 以下，大致呈线性分布，带宽为 100Hz 左右；在 1000Hz 以下呈对数增长。类似于临界带的划分，可以将语音频率划分成一系列三角形的滤波器序列，即 Mel 滤波器组，如图 4-6 所示。

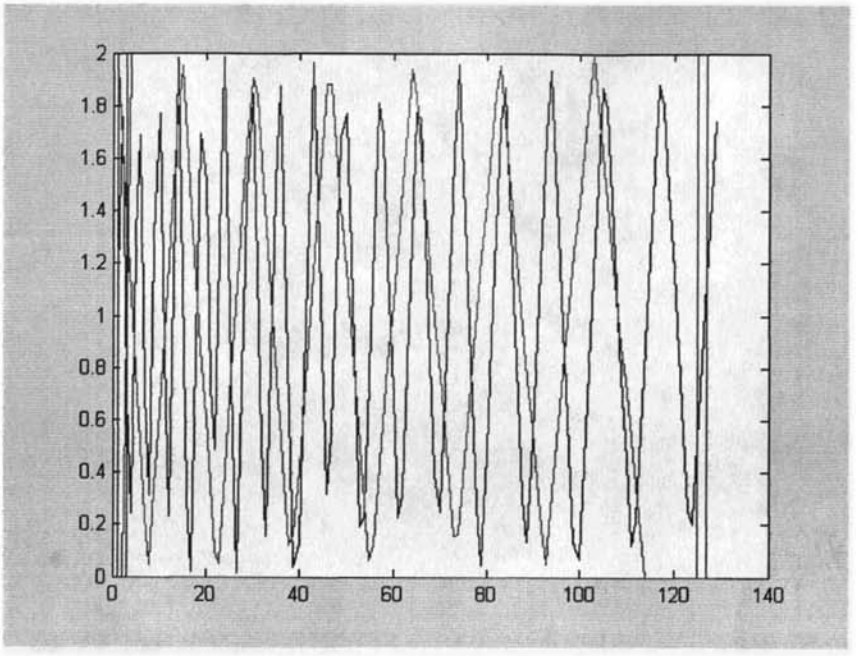

图 4-6 Mel 尺度滤波器组

MFCC 参数也是按帧计算的。首先要通过 FFT 得到该帧信号的功率谱 $S(n)$，转换为 Mel 频率下的功率谱。这需要在计算之前先在语音的频谱范围内设置若干个带通滤波器：

$$H_m(n), \quad m=0,1,\cdots,M-1, \quad n=0,1,\cdots,N/2-1 \tag{4-2}$$

M 为滤波器的个数，通常取 24，N 为一帧语音信号的点数，为了计算 FFT 的方便，通常取 N 为 256。滤波器在频域上为简单的三角形，其中心频率为 f_m，它们在 Mel 频率轴上是均匀分布的。在线性频率上，当 m 较小时，相邻的 f_m 间隔很小，随着 m 的增加，相邻的 f_m 间隔逐渐拉开。另外在频率较低的区域，f_m 和 f 之间有一段是线性的。带通滤波器的参数事先计算好，在计算 MFCC 参数时直接使用。

MFCC 的计算通常采用如下的流程：

（1）首先确定每一帧语音采样序列的点数，本课题取 $N=256$ 点。对每帧序列 $s(n)$ 进行预加重处理后再经过离散 FFT 变换，取模的平方得到离散功率谱 $S(n)$。

（2）计算 $S(n)$ 通过 M 个 $H_m(n)$ 后所得的功率值，即计算 $S(n)$ 和 $H_m(n)$ 在各离散频率点上乘积之和，得到 M 个参数 P_m，$m=0,1,\cdots,M-1$。

（3）计算 P_m 的自然对数，得到 L_m，$m=0,1,\cdots,M-1$。

（4）对 $L_0,L_1,\cdots,L_{m-1}$ 计算其离散余弦变换，得到 D_m，$m=0,1,\cdots,M-1$。

（5）舍去代表直流成分的 D_0，取 $D_1,D_2,\cdots,D_K$ 作为 MFCC 参数。此处 $K=12$。

图 4-7 是女声发音“打开”的原始语音序列及其 MFCC 参数序列的对比图，由图我们可以看出，MFCC 参数序列有效的保留了原语音信号的波形频率等特征，故本课题采用 MFCC 参数作为特征提取的特征矢量，以便于下一环节进行模式匹配的处理。

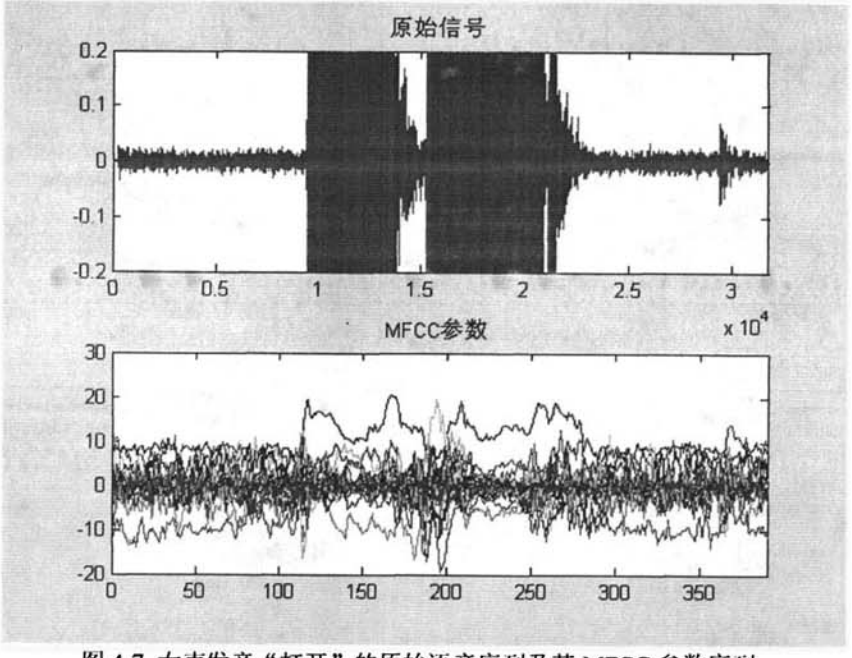

图 4-7 女声发音“打开”的原始语音序列及其 MFCC 参数序列

4.4 采用蚁群动态时间规划算法进行识别

语音信号经过前面几个模块的预处理、特征提取，再经过欧氏距离测度，就可以与已训练好的模板库进行匹配，从而得出识别结果。由于语音信号特征参数序列往往是不等长的，所以要寻找一个最佳的时间规正函数，使被测语音

模板的时间轴i，非线性地映射到参考模板的时间轴j，使总的累计失真量最小。传统的 DTW 算法在匹配参考模板与被测模板时通过局部优化的方法实现加权距离总和最小，本课题中，我们利用蚁群算法优化机制，结合传统的 DTW 算法，提出了一种新的基于蚁群算法的动态时间规划算法来搜索语音信号特征参数序列之间匹配的一条全局最优路径，进而以此衡量语音信号之间的相似度，从而期望得到最佳的识别结果。

在第 3 章的 3.5 节中，我们已经详细分析了蚁群动态时间规划算法的算法原理以及信息素更新机制。根据 4 个约束条件，蚂蚁从当前点到下一点所允许的集合如公式(3-12)，同时象 DTW 算法一样，我们把蚂蚁的活动范围约束在一个上边斜率为 2 下边斜率为 1/2 的平行四边形内，如图 4-8 所示。这样我们就可以保证蚂蚁在坐标轴上每一步只能走上、右、右上方向，而且必然最终到达终点$e=(M,N)$。从而我们可不必象基本蚁群算法那样设置一个禁忌表，只要对蚂蚁的路径做如上约束即可。

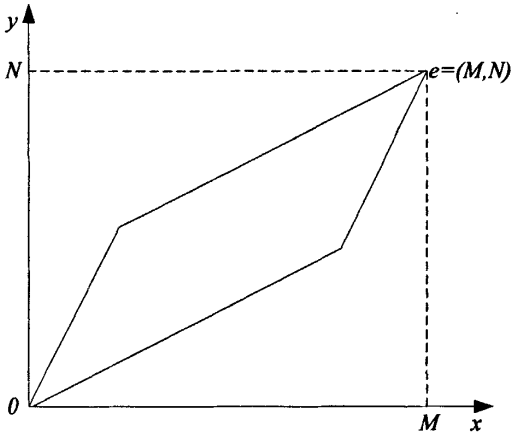

图 4-8 蚂蚁的路径范围

在本课题中，蚁群动态时间规划算法的系统参数设置如下：

窗宽r为 15；$\alpha=1$，$\beta=5$，信息素衰减因子$\rho=0.65$；初始信息素设为 1。

蚁群动态时间规划算法程序流程图如下所示：

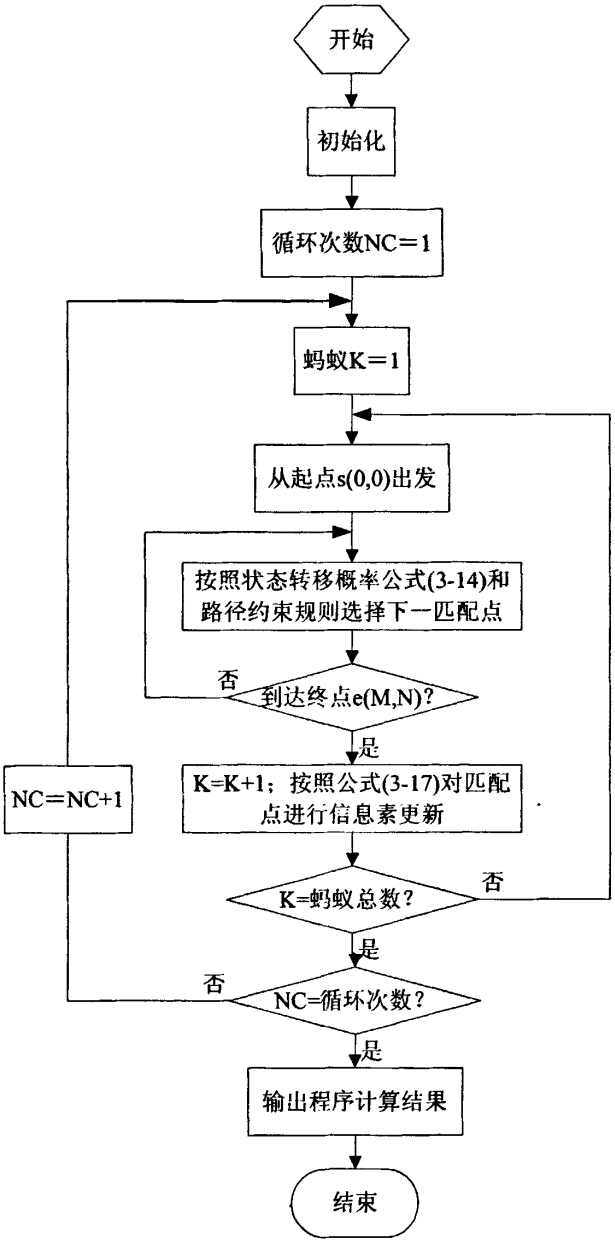

图 4-9 蚁群动态时间规划算法程序流程图

我们用同一女声但不同时间录取的发音“你好”分别做参考模板与被测模板做实验。分别取群体中蚂蚁个数为 5、10、15、20、25、30；循环次数为 5、10、15、20、25、30。则由实验结果得到表 4-1。表中蚂蚁个数为 k，循环次数

为 NC，D_k 为当蚂蚁个数为 k 时求得的匹配路径相对应的参考语音与测试语音之间的全局平均似然比失真，它体现了两模板之间语音信号的声学相似特性。

表 4-1 蚁群动态时间规划算法实验数据

NC	D_5	D_{10}	D_{15}	D_{20}	D_{25}	D_{30}
5	0.3157	0.2367	0.2253	0.2235	0.2132	0.2145
10	0.2468	0.2593	0.2255	0.2221	0.2127	0.2039
15	0.2589	0.2293	0.2033	0.2023	0.2018	0.2008
20	0.2436	0.2109	0.2032	0.1950	0.1937	0.1941
25	0.2439	0.2098	0.2091	0.1952	0.1934	0.1927
30	0.2431	0.2181	0.2015	0.1990	0.1942	0.1926

由表 4-1 可以看出，群体中蚂蚁数目 k 一定，随着循环次数 NC 的增加，或当循环次数 NC 一定，随着蚂蚁数目 k 的增加，匹配路径的平均累计失真 D 总体呈下降趋势，可见其寻优过程是有效的。当 k 达到 20，NC 达到 20 时，匹配路径的失真度随着 k 与 NC 的增加，D 变化趋势已不明显，且考虑运算量与运算速度的问题，实验数据表明 k=20，NC=20 寻到的路径为匹配最优路径，即最大程度的体现语音信号的相似性。所以本课题取蚂蚁 k=20，NC=20。

当循环次数 NC =20 时，蚁群动态时间规划算法在寻找最优匹配路径过程中的全局平均似然比失真 D_k 和蚂蚁数 K 之间的关系如图 4-10 所示。

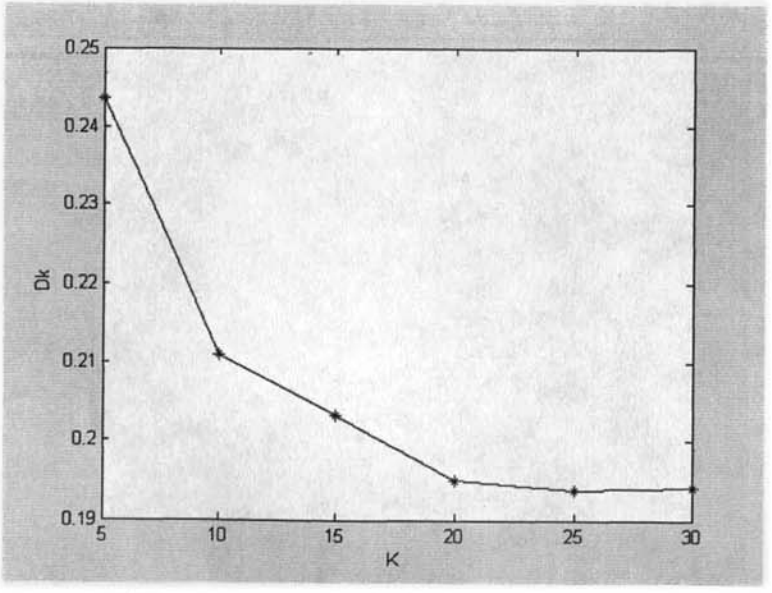

图 4-10 循环次数 NC=20 时蚁群动态时间规划算法的最优化过程

图 4-10 表明，当蚂蚁数 k 达到 20 失真 D_k 基本保持不变。根据以上分析，可知本课题提出的蚁群动态时间规划算法是可行的，且本文提出的状态转移概率方程与信息素更新原则，在寻求匹配最优路径的结果是有效的。

4.5 实验结果

实验所采用的数据都是通过麦克风由计算机的声卡录音得到的，采样频率是 8000Hz，语音数据在不同时间录音。语音数据包括孤立词语音数据与连续语音数据，其中孤立词语音数据为单个字放、停、前、后、开、关；连续语音数据为播放、停止、前进、倒退、打开、关闭。本系统是为车载信息平台中的语音控制媒体播放系统设计，因为使用这个系统的并不仅是特定人，所以以上语音数据又有特定人发音和非特定人发音之分。特定人发音的数据来自一个女声发音的 20 组语音，也就是每个指令词有 20 个不同的发音，总共是 240 个语音数据；非特定人发音的数据来自两男两女的发音，建立了一个 240×4 的语音数据库。

4.5.1 对孤立词语音识别的实验

本实验中采用了特定人与非特定人两个孤立词语音数据库分别测试了基于蚁群动态时间规划算法以及 DTW 算法的语音系统的性能。

（1）特定人汉语孤立词语音识别实验

实验数据为一个女声对单个字放、停、前、后、开、关的发音，共 120 个。每次进行 4 组数据的测试，其中有 2 个为同一字的发音，另外 2 个为除这个发音外在令 5 个发音中的任意组合。每个单字都测试到，而且为了防止误差每个单字都测试了 50 遍。

（2）非特定人汉语孤立词语音识别实验

实验数据为两男两女对单个字放、停、前、后、开、关的发音，共 480 个。每次进行 4 组数据的测试，其中有 2 个为不同人对同一字的发音，另外 2 个为除这个发音外在令 5 个发音中的任意组合。每个单字都测试到，而且为了防止误差每个单字都测试了 100 遍。

测试数据如表 4-2 所示，为了对比，我们把基于蚁群动态时间规划算法以及 DTW 算法的测试数据列在一起。

表 4-2 两种算法孤立词识别率的比较

识别类型	蚁群动态时间规划算法	DTW 算法
特定人识别率（平均）	100%	100%
非特定人识别率（平均）	98．83%	98．17%

从表 4-2 可以看出，两种算法对于指令词的识别较理想，特别是对于特定人单个命令字识别率都达到了 100%，这主要是因为词汇本身发音较短，且这 6 个词的发音都各有特征不易混淆，模板之间有很强的区别性，易于进行区分。在非特定人识别率的试验中，蚁群动态时间规划算法比 DTW 算法的识别率稍有提高。虽然说差距不大，但对于简单的孤立词识别系统中，这种差距已经是一种比较大的优势。

4.5.2 对连续语音识别的实验

本实验中采用了特定人与非特定人两个连续语音数据库分别测试了基于蚁群动态时间规划算法以及 DTW 算法的语音系统的性能。

（1）特定人汉语连续语音识别实验

实验数据为一个女声对指令词播放、停止、前进、倒退、打开、关闭的发音，共 120 个。每次进行 4 组数据的测试，其中有 2 个为同一词的发音，另外 2 个为除这个发音外在令 5 个发音中的任意组合。每个指令词都测试到，而且为了防止误差每个指令词都测试了 50 遍。

（2）非特定人汉语连续语音识别实验

实验数据为两男两女对指令词播放、停止、前进、倒退、打开、关闭的发音，共 480 个。每次进行 4 组数据的测试，其中有 2 个为不同人对同一词的发音，另外 2 个为除这个发音外在令 5 个发音中的任意组合。每个指令词都测试到，而且为了防止误差每个指令词都测试了 100 遍。

测试数据如表 4-3 所示，同样为了对比，我们把基于蚁群动态时间规划算法以及 DTW 算法的测试数据列在一起。

表 4-3 两种算法连续语音识别率的比较

识别类型	蚁群动态时间规划算法	DTW 算法
特定人识别率（平均）	99．33%	98．67%
非特定人识别率（平均）	96．17%	94．83%

从表 4-3 可以看出，对于连续语音识别，其识别率比孤立词语音识别率稍有下降。DTW 算法在非特定人的连续语音实验环节其识别率更是低于 95%，当然识别率的问题不仅是由模式匹配的算法决定的，也受系统前面的预处理、特征提取的影响。单纯就这个环节来说，蚁群动态时间规划算法又表现出比 DTW 更好的性能，特别是在非特定人的连续语音实验部分表现的更为突出。情况越复杂，蚁群动态时间规划算法的表现就越优越，这也是因为蚁群动态时间规划算法在计算全局平均失真时，每次要进行 20 遍迭代，每次迭代都要不停的进行信息素更新，最终得到的最优路径比 DTW 算法可以更准确的表示语音信号之间的相似性，体现了此算法的优越性。

4.5.3 分析讨论

在识别率方面，蚁群动态时间规划算法与 DTW 算法对于指令词的识别较理想，特别是对于特定人单个命令字识别率都达到了 100%，在非特定人识别率的试验中，蚁群动态时间规划算法比 DTW 算法的识别率稍有提高。虽然说差距不大，但对于简单的孤立词识别系统中，这种差距已经是一种比较大的优势。对于连续语音识别，其识别率比孤立词语音识别率稍有下降。DTW 算法在非特定人的连续语音实验环节其识别率更是低于 95%。单纯就这个环节来说，蚁群动态时间规划算法又表现出比 DTW 更好的性能，特别是在非特定人的连续语音实验部分表现的更为突出。情况越复杂，蚁群动态时间规划算法的表现就越优越，这也是因为蚁群动态时间规划算法在计算全局平均失真时，每次要进行 20 遍迭代，每次迭代都要不停的进行信息素更新，最终得到的最优路径比 DTW 算法可以更准确的表示语音信号之间的相似性，体现了此算法的优越性。另一方面，蚁群动态时间规划算法还有潜力可挖，采用它的语音识别系统在复杂情况下还有望进一步提高识别率。我们需要做更多的实验进一步深入研究相应的蚁群算法的参数取值，同时改进蚁群算法的信息素更新机制，寻求更为合理可

靠的蚁群算法模型形式。

在实验过程中，我们也注意到，当改变实验条件时，语音识别系统的环境适应性还不够。主要体现在对环境依赖性强，即待测语音和训练语音要在相同的环境中获得，否则系统识别率会下降。例如我们采用不同的环境又做了特定人汉语连续语音识别实验。同样的一个女声发音，一部分在实验室采集得到，另一部分通过笔记本在学校的广场上采集得到。这次的实验结果就和在同一环境下得到的不一致了。实验结果如下表 4-4 所示：

表 4-4 不同环境下连续语音特定人识别率的比较

识别类型	蚁群动态时间规划算法	DTW 算法
识别率（平均）	94．47%	91．31%

当然这和采用的测度估计算法无关，因为无论是蚁群动态时间算法还是 DTW 算法都有这种情况，问题并不出在这块。这有关于系统的适应性以及抗噪性能，这方面的改进将在下一步的工作中完成。

虽然蚁群动态时间规划算法识别率要比传统的 DTW 算法更高，但是其识别速度却要比 DTW 算法稍慢。在对非特定人汉语连续语音识别的实验中，识别出一个待测模式大概需要 2 秒左右（C 语言环境，P4 2.0 的 CPU，512M 的 RAM），而普通的 DTW 算法大概需要 1 秒左右的时间。这可以从时间复杂度上进行解释：DTW 算法的时间复杂度为 $O(M\times N)$，其中 M、N 分别为参考模板和待测模板的特征矢量序列长度；蚁群动态时间规划算法的时间复杂度为 $O(N_c\times(M^2+N^2)\times K)$，其中 M、N 分别为参考模板和待测模板的特征矢量序列长度，N_c 和 K 分别为迭代次数以及蚂蚁总数。从时间复杂度上看，蚁群动态时间规划算法相比 DTW 算法稍慢，但这种差距也只是线性的，并没有达到指数级。故我们可以得出，蚁群动态时间规划算法是完全可行的，速度上的这点细微差距在实际应用时可以采用相对快速的处理器来弥补；而且我们还可以对蚁群算法进行改进，使其更有效率，这一工作有待于进一步研究。

在空间复杂度上，普通 DTW 算法需要 $2\times M\times N$ 的空间存储帧匹配距离矩阵和累积距离矩阵，而蚁群动态时间规划算法的空间复杂度为 $O(M^2+N^2)+O(\sqrt{M^2+N^2}\times K)$，但在这里 $K<<\sqrt{M^2+N^2}$，因此整个算法的空间复杂度为 $O(M^2+N^2)$。这和 DTW 算法差不多，可见数据存储上基本蚁群算法的空间复杂度是简单的，易于实现，在语音识别系统中，蚁群动态时间规划

算法不失为一种很好的方法。

通过以上对识别率、系统适应性、运算速度、算法时间复杂度以及空间复杂度的分析，我们可以得出蚁群动态时间规划算法是一种能够在语音识别系统中替代 DTW 算法的智能优化算法，并且这个算法在语音识别系统中还有很大的潜力可挖掘，通过优化选择蚁群算法的参数取值、改进蚁群算法的信息素更新机制能使其在语音识别系统中的性能得到进一步提升，这些任务将在以后的工作中进一步完成。

第 5 章　总结与展望

5.1　课题总结

本课题通过对语音信号的端点检测、特征提取以及识别算法等方面的研究分析，利用蚁群算法优化机制，结合传统的 DTW 算法，设计了一种基于蚁群算法进行动态时间规划的语音识别系统，初步达到了实用性的要求，本论文主要工作如下：

（1）根据语音识别系统的特点与要求，提出了从预处理、端点检测、特征提取到匹配识别的完整思路，给出了详细的方案与方法。

（2）通过对语音信号处理和基本蚁群算法的理论与方法的系统研究，将蚁群算法应用于语音识别，论证了其可行性和适用性。

（3）利用蚁群算法优化机制，结合传统的 DTW 算法，设计出了蚁群动态时间规划算法。在运算速度差不多的前提下，有效的提高了系统识别率。

（4）对新的语音识别系统进行仿真测试。为了有效地验证算法的可行性与软件性能，算法测试所用的数据是实际的语音信号通过麦克风，经计算机声卡采集而获得的，以计算机文件的形式存储为语音数据文件。

5.2　工作展望

语音信号处理及蚁群算法均是目前研究的热点，本文工作只是在将蚁群算法引入语音信号识别方面的初步尝试，其中还有许多理论和应用问题需要继续深入探讨。建议下一步的工作可以展开以下研究：

（1）语音识别系统的环境适应性还不够，主要体现在对环境依赖性强，即待测语音和训练语音要在相同的环境中获得，否则系统识别率会下降。未来可以考虑进一步改进端点检测算法以增强系统的噪声鲁棒性。

（2）根据语音识别系统的特点与要求，进一步深入研究相应的蚁群算法的参数取值，改进蚁群算法的信息素更新机制，寻求更为合理可靠的蚁群算法模型形式。

（3）关于识别速度问题：由于本系统采用的是蚁群动态时间规划算法，在

词汇量不大的情况下，系统识别速度完全可以满足实时要求。但是如果将词汇量扩大很多，速度就将会下降。在这种情况下，为了提高识别速度，我们可以对蚁群算法进行改进，使其更有效率。

（4）利用虚拟仪器等现代计算机技术，研发专用的语音识别软硬件系统，增强语音识别的实用性，争取早日移植到车载信息平台中。

有所创新，才能有所发展。任何一个学科的发展都是一个提出问题、解决问题的过程，更是几代甚至几十代科学家付出大量心血、不断探索、勇于实践的过程。语音信号的处理与识别学科的发展更是如此，其中所面临的许多问题都需要人们不断的去寻求新的解决方法。在今后的研究工作中，需要注意吸取其它学科的理论知识，勇于挑战科技前沿，使语音信号处理与识别的研究工作再上一个新的台阶。

参考文献

[1] 韩纪庆，张磊，郑铁然.语音信号处理[M].北京：清华大学出版社，2005

[2] 蔡莲红，黄德智，蔡锐.现代语音技术基础与应用[M].北京：清华大学出版社，2004

[3] M. Dorigo, G. DiCaro.The Ant Colony Optimization [A].NewMeta-Heuristic, Proceedings of the Congress on Evolutionary Computation [C], London,UK, 1999:11~32

[4] 陈峻，沈洁，秦玲.蚁群算法求解连续空间优化问题的一种方法[J].软件学报，2002.13 (12):2317-2323

[5] 马良.基于蚂蚁算法的函数优化[J].控制与决策，2002,17(增刊): 719~726.

[6] 易克初，田斌，付强.语音信号处理[M].北京：国防工业出版社, 2000

[7] 何强，何英.Matlab 扩展编程[M].北京：清华大学出版社，2002

[8] 姚天任.数字语音处理.湖北：华中科技大学出版社，2002

[9] General Aspects of Digital Transmission System, Coding of Speech At 8kbit/s Using Conjugate-sturcture Algebraie-code-excited Linear-predietion (CS-ACELP).JTU-T Recommendation G.729

[10] 周德俊，杨莉.G.729 语音压缩编码及其 DSP 实现.通信技术，2001(4)

[11] 杨行峻，迟惠生.语音信号数字处理[M]. 北京：电子工业出版杜，2003

[12] 胡航.语音信号处理(第2版) .哈尔滨：哈尔滨工业人学出版社，2000

[13] M.H. Savoji.A Robust Algorithm for Accurate End Pointing of Speech.Speech Communication, 1989, 8(2):45~60

[14] R.Bhiksha, S.Rita.Classifier-based Non-linear Projection for Adaptive End Pointing of Continuous Speech.Computer Speech&Language, 2003,17(l):5~26

[15] 聂敏.语音识别及其关键技术[J].微波与卫星通信，1999,4:53~56

[16] C. Lee, D. hyun, C. Nadeu.Optimizing feature extraction for speech recognition [J].IEEE Transactions on Speech and Audio Processing, 2003, 11(l):80~86

[17] 王炳锡.语音编码[M].西安：西安电子科技大学出版社，2002

[18] 赵力.语音信号处理.北京：机械工业出版社，2003

[19] 徐宵鹏，吴及.孤立词语音识别算法性能研究与改进.计算机工程与应用，2001,21:144~146

[20] L. Rabiner, B.H. Junag.Fundamentals of Speech Recognition.PTR Prentice Hall, 1993

[21] Dupont, Stephane, Cheboub, Leila.Fast seaker adaptation of artificial neural networks for automatic speech recognition.IEEE Intemational Coneferneeon Acoustics, Speech and Signal Processing-Proceedings, 2000

[22] A. Colorni, M. Dorigo, V. Maniezzo.Distributed optimization by ant colonies.Proceedings of the 1st European Conference on Artificial Life, 1991:134~142

[23] M. Dorigo.Optimization, learning and natural algorithms.Ph.D Thesis, Department of Electronics, Politecnico di Milano, Italy, 1992

[24] M. Dorigo, V. Maniezzo, A. Colorni.Ant system: optimization by a colony of cooperating agents.IEEE Transaction on Systems,Man,and Cybernetics-Part B, 1996,26(1):29~41

[25] M. Dorigo, V. Maniezzo, A. Colorni.Positive feedback as a search strategy.Technical Report 91~016, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1991

[26] V. Maniezzo, A. Colorni, M. Dorigo.The ant system applied to the quadratic assignment problem.Technical Report IRIDIA/94-28, IRIDIA, Universite Libre de Bruxelles, Belgium, 1994

[27] L.M. Gambardella, E.D. Taillard, M. Dorigo.Ant colonies for the QAP.Technical Report IDSIA-4-97, IDSIA, Lugano, Switzerland, 1997

[28] L.M. Gambardella, E.D. Taillard, M. Dorigo.Ant colonies for the quadratic assignment problem.Journal of the Operational Research Society, 1999,50(2):167~176

[29] A. Colorni, M. Dorigo, V. Maniezzo.Ant system for job-shop scheduling.Belgian J.Oper.Res.Statist.Comput.Sci, 1994,34:39~53

[30] D. Costa, A. Hertz.Ants can colour graphs.Journal of the Operational Research Society, 1997,48(3):295~305

[31] S.H. Ahn, S.G. Lee, T.C. Chung.Modified ant colony system for coloring graphs.Proceedings of the 2003 Joint Conference of the 4th International Conference on Information,Communication and Signal Processing and the 4th Pacific Rim Conference on Multimedia, 2003,3:1849~1853

[32] M. Dorigo, V. Maniezzo, A. Colorni.Positive Feedback as a Search Strategy.Technical Report 96~106

[33] 段海滨，王道波.一种快速全局优化的改进蚁群算法及仿真.信息与控制，2004,33(2):241~244

[34] 詹士昌，徐婕，吴俊.蚁群算法中有关算法参数的最优选择.科技通报，2003,9(5):381~386.

[35] H.M. Botee, E. Bonabeau.Evolving ant colony optimization.Advances in Complex Systems,

1998,1(2):149~159

[36] Z.W. Wanda, O. Tokunbo.Formant and Pitch Detection Using Time-frequency Distribution.International Journal of SpeeehTeehnolog, 1999,3(l):35~49

[37] 段海滨.蚁群算法及其在高性能电动仿真转台参数优化中的应用研究.南京：南京航空航天大学博士学位论文，2005

[38] J. Makhoul, A. Gray.Linear Prediction of Speech.Springer-Verlay, 1996

[39] W.H. Shin.Speech/non-speech Classification Using Multiple Features for Roust Endpoint Detection.Proceedings of IEEE ICASSP, Istanbul, 2000,3:1399~1402

[40] C.S. Huang, H.C. Wang.Bandwidth-adjusted LPC Analysis for Robust Speech Recognition.Pattern Recognition Letters, 2003,24(9):1593~1597

[41] R. Bennetl, A. Syndal, S. Greenspan.Applied speech technology.USA Florida: CPC press, 1995

[42] 江太辉.基于 DTW 算法的语音识别电话系统.电声技术，2005,8:31~34

[43] N.T lay, F.W. Say, D. Silva, ete.Speech Emotion Recognition Using Hidden Makrov Models.Speech Communication, 2003,41(4):603~623

[44] 王玥，陶洪久.蚁群优化算法在 TSP 中的应用.武汉理工大学学报信息与管理工程版，2006,28(11):24~26

[45] 段海滨，王道波，朱家强等.蚁群算法理论及应用研究的进展.控制与决策，2004,19(12):1321~1326

致 谢

我衷心感谢导师黄涛副教授，本文是在他的悉心指导下完成的。黄涛老师给予了我细致的指导、热情的鼓励和全面锻炼的诸多机会，使得我的学位论文得以顺利完成。黄涛老师学识渊博、思维敏捷、思路宽广、治学严谨，他不仅授予我丰富的专业知识，还教我懂得了许多做人的道理，在此我由衷地感谢黄涛老师。

另外，武汉理工大学智能信息系统研究所的廖传书副教授和卢骆先副教授。两位老师在我 3 年的研究生生涯中给予我诸多指导和教诲，这对于我人生的影响无疑是积极而深远的，在这里衷心的感谢两位老师。

我还要感谢同实验室语音组的贺宽、陈鹏飞同学以及实验室的其他同学，他们都给予我许多无私的帮助，也谢谢他们。

我要感谢我的父母、女友，他们的支持和鼓励是我前进的动力，对他们的感谢很难用言语来表达。

感谢所有帮助和关怀过我的人，谢谢你们！

最后，衷心的感谢在百忙之中评阅论文和参加答辩的各位专家、教授！

附　录 攻读硕士学位期间发表的学术论文

[1] 黄涛，肖宜. 蚁群算法在语音识别中的应用研究. 武汉理工大学学报（信息与管理工程版）. 第 29 卷，第 12 期，2007 年 12 月

[2] 肖宜. 语音识别中双门限端点检测算法的研究. 中国科技论文在线. 2008 年 4 月