河北大学

硕士学位论文

基于谱熵的语音端点检测算法的研究

姓名：张翠改

申请学位级别：硕士

专业：通信与信息系统

指导教师：龙海南

20100501

摘 要

在当代各种各样的通信系统中，语音通信业务一直是必不可少的，如何在一段语音信号中将有效的语音和无效的噪声分离，多年来，这个问题一直受到众多学者的关注。

所谓的语音端点检测就是指从一段语音信号中准确的找出语音信号的起始点和结束点，它的目的是为了使有效的语音信号和无用的噪声信号得以分离。它作为一种语音信号预处理技术，在实际应用中起着非常重要的作用，由于消除了背景噪声的影响，语音处理的精度得以提高。另外它的应用范围十分广泛，在语音识别、语音增强、语音编码、回声抵消等系统中都是必不可少的环节。

经过许多专家学者多年的研究，已经提出了许多语音端点检测的方法，这些方法大体上分为两大类。一类是基于阈值的方法，这种方法就是根据语音信号和噪声信号的不同特征，提取每一段语音信号的特征，然后把这些特征值与设定的阈值进行比较，从而达到语音端点检测的目的，最传统的一些算法例如短时能量，过零率等就属于这一类。另一类方法是基于模式识别的方法，需要估计语音信号和噪声信号的模型参数来进行检测。基于阈值的检测方法原理简单，运算方便，从而被人们广泛使用，但是由于在信噪比降低的情况下，语音信号容易被噪声淹没，所以它的检测效果会变的很差，而基于模式识别的方法又由于自身复杂度高，运算量大，很难被人们应用到实时语音信号系统中去。

本课题将子带自适应选择谱熵检测算法以及自适应滤波技术相结合，达到先滤波再检测的目的，和其他语音端点检测算法相比，该算法具有较高的鲁棒性，在信噪比降低的条件下能比较准确的进行语音端点检测。

在本课题中应用 MATLAB 软件对本文中选用的语音端点检测算法进行仿真，表明该算法具有较好的鲁棒性。选用 TMS320VC5416 DSP 芯片作为核心处理器以及 TLV320AIC23 作为 Codec 芯片来研究语音端点检测的硬件系统。

关键字 语音端点检测 谱熵 子带自适应选择 自适应滤波

Abstract

Nowadays, in various comminication systems, speech communication service is of the most important ones. These years, many scholars have focused on how to separate the useful speech from the unuseful noise.

Detect the starting point and ending point from a section of speech signal, this procedure is called speech endpoint detection, in this way we can separate the useful speech signal and the useless noise. As a preprocessing technology of speech signal, speech endpoint detection plays an important part in practical applications. For erasing the effects of background noise, the accurate of speech processing can be improved. It's widely used in speech recognition, speech amplification, speech coding, echo cancellation.

Through several years of research, there are many speech endpoint detection approaches been proposed which can be divided into two classes. One is based on threshold value, as speech signal and noise signal have different features, we can extract the features of each speech section, and compare the value of the features to the threshold value, then obtain the purpose of speech endpoint detection. The earliest ones include short-time energy, zero-crossing rate and so on.The other one is based on model recognition, in this way we need to estimate the model parameter of each speech signal and noise signal for detecting. The princapal based on threshold value is simple and ease to calculate, so it is widely used. However, if the signal-to-noise ratio (SNR) is too low, the speech signal can be submerged in the noise, and its detecting effects can be worse. And the approaches based on model recognition are far more likely to be used in real-time speech signal system for its complication and huge compution.

In this paper, we combined the algorithm of adaptive band-partitioning spectral entropy with adaptive filter.To obtain the purpose of reducing noise first and then detection, compared to other approaches, this one is more robust and can operate effectively under low SNR environment.

In the paper, we used matlab to simulate the algorithm,and the result demonstrates its robustness.moreover，we choose TMS320VC5416 as processor and TLV320AIC23 as codec chip to study the hardware system of the speech endpoint detection.

Keywords Speech Endpoint Detection Spectral Entropy Adaptive Band-partitioning Adaptive Filter

河北大学

学位论文独创性声明

本人郑重声明： 所呈交的学位论文，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，论文中不包含其他人已经发表或撰写的研究成果，也不包含为获得河北大学或其他教育机构的学位或证书所使用过的材料。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致谢。

作者签名：张翠改　　日期：2010 年 6 月 4 日

学位论文使用授权声明

本人完全了解河北大学有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，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、缩印或其他复制手段保存论文。

本学位论文属于

1、保密 □ ，在_____年_____月_____日解密后适用本授权声明。

2、不保密 ☑。

（ 请在以上相应方格内打“ √ ” ）

保护知识产权声明

本人为申请河北大学学位所提交的题目为（基于谱熵的语音端点检测算法的研究）的学位论文，是我个人在导师（龙海南）指导并与导师合作下取得的研究成果，研究工作及取得的研究成果是在河北大学所提供的研究经费及导师的研究经费资助下完成的。本人完全了解并严格遵守中华人民共和国为保护知识产权所制定的各项法律、行政法规以及河北大学的相关规定。

本人声明如下：本论文的成果归河北大学所有，未经征得指导教师和河北大学的书面同意和授权，本人保证不以任何形式公开和传播科研成果和科研工作内容。如果违反本声明，本人愿意承担相应法律责任。

声明人：张翠改 日期：2010年6月4日

作者签名：张翠改 日期：2010年6月4日

导师签名：龙海南 日期：2010年6月4日

第 1 章 引 言

1.1 语音端点检测的背景和意义

1.1.1 语音端点检测的背景

语言是人类特有的一种功能，是人们进行情感表达和思想交流的最重要的途径，语言的声学表现形式就是语音，语音是人类最自然、最有效、最直接、最重要的信息载体，换句话说，它是人类交流中最简单最必不可少的一种工具。人类可以利用多种手段获得外界信息，其中最重要的，表达也最准确的手段只有语言、图像和文字三种，众所周知，利用语言来传递信息明显要比利用视觉和文字来传递信息有效的多[1]。时至今日，在信息高速发达的环境下，语音仍然是人类进行信息交换的最主要的工具，所以，长久以来，人类对语音的关注热情非但从未降低，而且随着科学技术的发展以及对语音处理技术的提高，人类对语音的研究更加深入。

1.1.2 语音端点检测的意义

随着现代科学和计算机技术的快速发展，语音识别成为信息科学中一个重要的研究领域。所谓的语音识别，又称自动语音识别（Automatic Speech Recognition，简称 ASR），其研究目标是如何使机器能够准确地听出人的语音及其内容，以便控制其他设备来满足人类的各种需要或者实现人机自然交流。语音识别的研究与声学、发声机理和听觉机理、语音学、语言学、信号处理理论、概率论和信息论、模式识别理论、最优化理论、人工智能、计算机科学等学科有着紧密的联系。语音的端点检测是语音识别系统中预处理技术的一部分，是该领域内的一个关键性问题，因而随着语音识别技术的快速发展，也相应得到了研究学者的普遍关注。

所谓的语音端点检测就是指从一段语音信号中正确检测出语音的起始点和结束点，从而将有效语音信号检测出来。它是语音信号预处理技术的一部分，有着广泛的应用范围，它在语音识别、语音通信、语音增强、说话人识别中起着非常重要的作用。有效的端点检测技术不仅能减少系统的处理时间，提高系统的处理实时性，而且能排除无声段的噪声干扰，从而使后续工作的性能得到较大的提高；相反不准确的端点检

测可以引起识别率的下降和计算量的增加。所以研究端点检测技术对于语音信号处理可起到关键作用。

1.2 语音端点检测技术的研究现状

对语音端点检测的研究最早可以追溯到上个世纪 50 年代。当时是在一个实时语音翻译系统中，为了解决语音段和无语音段的检测问题而提出来的，并把该算法命名为VAD（voice Activity Detection），指的是将语音段和无语音段分开的处理过程[2]。

传统的语音端点检测算法都是针对较为安静的实验室环境进行的，近年来，各种噪声环境下的语音端点检测也被人们纳入研究之列。经过几十年的研究，许多端点检测算法被相继提出，这些算法主要被分为两类，一类是基于特征的端点检测方法，一类是基于模型的端点检测算法。前者包括基于短时能量[3]、过零率（zero-crossing-rate，ZCR）[4]、熵[5]、LPC倒谱距离的算法[6]等等，基于模型的检测算法最常见的是基于隐马尔可夫模型[7], 支持向量机[8]和神经网络[9]的算法等等。

理想的语音端点检测算法应当能够满足可靠性、鲁棒性、精确性、自适应性、实时性和对噪声特征无需先验知识等。现有的算法存在的主要问题有两点：一个是在具有较强噪声的环境下，语音端点检测往往存在着大量的误判，不利于后续的处理过程；另外一个是在高噪声的环境下不能有效的检测出语音信号段，造成了有效信息的丢失，以上两个问题也得到了广大研究者的重视。近几年来，研究者们经过了不懈的努力，提出了各种区别语音和噪声的特征参数，用来提高算法的抗噪声性能，或是将几种特征组合成一个新的特征参数来进行端点检测，而对语音端点的判决也由原来的单一门限发展到多门限以至于自适应门限，使得算法精度不断得到提高，但是如何在噪声环境下设计一种鲁棒的端点检测算法仍是一个亟待解决的棘手问题。

1.3 DSP 技术的发展与应用

DSP既是Digital Signal Processing 的缩写，也是Digital Signal Processor的缩写。前者指数字信号处理的理论和方法，后者是指用于数字信号处理的可编程微处理器。随着科学技术的发展，DSP芯片已经成为通信、计算机、网络、工业自动控制以及日常家用电器等电子产品中不可缺少的重要器件。

DSP 技术的发展经历了三个阶段。

70年代是数字信息处理技术的理论研究阶段，其代表是《Digital Signal Processing》，

由A. V. Oppenheimh 和R. V. Schafer 编写。数字信号处理技术的出现，成为分析实际现象的有力工具，当时的DSP系统由分离元件组成。

80 年代，DSP技术的发展得到了推动，因为人们对信息处理速度的要求不断提高。DSP此时已经成为一种专门处理信号的微处理器，它的优点是能快速对输入输出数据进行处理。1982年，第一代低成本高性能的DSP由美国TI公司研制成功，这样使得DSP开始在工业中得到广泛使用。

从90年代起，DSP 技术飞速发展，使得DSP芯片的性能得到不断提高。由DSP和外围电路等设计完成特定功能的芯片的产生，更是加速了DSP的发展。以数字信号处理的理论和算法为基础的DSP芯片，专门完成各种实时数字信息处理运算。当前的DSP多数基于RISC结构以及采用CMOS工艺的DSP器件得到广泛推广，这使得它成为大容量DSP系统的主流器件[10]。

随着计算机科学技术、信号处理技术等一些信息技术的的高速发展，通信行业以及工业控制等行业中需要处理越来越多的数据信息，同时它们也对处理信号的准确性和实时性的要求越来越高。所以此时高性能DSP的出现弥补了低性能DSP的不足，满足了人们的要求，并且DSP价格的降低，也使得DSP芯片越来越受到人们的青睐。

由于DSP有着非常快速的运算能力，有着极高的性价比，所以在当代社会中，DSP的应用领域越来越广泛。随着技术的不断发展，DSP的性能会越来越高，在未来世界中，DSP器件的应用将会涉及到人们生活的方方面面。

1.4 本文的主要内容及结构安排

1.4.1 主要内容

本课题提出了一种先应用自适应滤波技术进行语音降噪，然后应用子带自适应选择功率谱熵的语音端点检测算法。文中的工作主要包括以下两个方面：

- 用 MATLAB 软件对语音检测算法进行仿真

该算法应用功率谱概率密度函数来构建谱熵值，并应用子带自适应选择方法以及自适应滤波技术来实现的语音端点检测。

由于语音信号是功率信号而非能量信号，在该算法中采用功率谱密度而不用直接变换所得的频谱来构造谱熵函数，这样只要语音的分布情况不变，谱熵值不会随信号幅度的大小而变化，由此可以看出谱熵对于噪声有一定的鲁棒性。

子带自适应选择功率谱熵法是一种新的端点检测方法，它的思想是将一帧语音分成若干个子带，再用谱熵法进行运算，子带的个数可以自适应选择。

自适应滤波方法是近 40 年以来发展起来的一种最佳滤波方法。是由 Widrow B. 等人于 1967 年在维纳滤波、卡尔曼滤波等线性滤波基础上发展起来的一种最佳滤波方法。

本课题把两者进行结合得到自适应滤波的子带自适应选择谱熵法，既达到了良好降噪的目的，又能用稳健性好的特征参数进行语音端点检测，从而从两个方面来提高端点检测的准确性。

- 语音端点检测的硬件研究

由于 DSP 处理器的处理速度快，运算能力强大等多种优点，近年来经常被用于语音编码和通信应用方面。TLV320AIC23 是 TI 公司的一款高性能、集成有模拟功能的立体声音频 Codec 芯片，它通常被用于可移动的数字音频播放和录音的模拟输入输出等应用系统中，本文中采用 TMS320VC5416 作为主处理器，以及选用 TLV320AIC23 作为 Codec 芯片，以此来研究语音端点检测的硬件系统构成。

1.4.2 结构安排

第一章 引言。详细介绍了语音端点检测的背景、意义以及研究现状，随后介绍了 DSP 技术的发展与应用。最后给出了本文的主要内容及结构安排。

第二章 主要介绍了语音信号和噪声信号的基本特点、语音信号的基础处理知识。主要知识有短时分析技术、语音信号的预加重、分帧和加窗以及语音信号的时域分析和频域分析等。

第三章 对传统的语音端点检测算法进行了分析和研究，主要有短时能量、短时过零率、短时相关函数、倒谱距离检测。

第四章 详细介绍了基于语音信号的功率谱构建的语音熵的端点检测算法，研究并改进了子带自适应选择的功率谱熵的语音端点检测，并对该算法进行仿真与分析。

第五章 介绍了自适应滤波技术的原理、基本结构以及发展情况，详细给出了 LMS 算法的推导过程，并对该算法进行了仿真。

第六章 将子带自适应选择的功率谱熵的语音端点检测与 LMS 自适应滤波技术相结合，给出了算法过程和仿真结果。

第七章 简单介绍了语音端点检测系统的硬件设计，对 VC5416 和 AICA23 芯片的功能和特点进行了简单介绍，并给出了其连接图。

第八章 结束语。对本文进行了总结以及提出了下一步将要进行的工作。

第 2 章 语音端点检测技术基础理论

语音是由一连串的信号所组成，它是组成语言的声音，是一种具有声学物理特征的物质。语音的产生是人在说话的过程中，由声门处的气流冲击声带而产生的振动，然后通过声道响应所形成。语音信号的频谱范围主要集中在 300Hz～3400Hz，通常电话中语音的频率范围大约为 300～3400Hz左右，取样率一般取 8KHz，但在实际语音信号处理中，取样率通常取 7～10KHz。在不清楚信号带宽的情况下，取样前应接入反混叠滤波器（低通滤波器），使其带宽限制在某个范围内，否则如果取样率不满足取样定理，则会产生频谱混叠，此时语音信号中的高频成分将产生失真[1]。

2.1 语音和噪声的特性

2.1.1 语音特性

语言的声学表现就是语音，语音信号的特性就是指语言的声学特性，主要包括以下四个方面。

（1）短时平稳性[11]

语音信号是一个随着时间变化的非平稳的随机过程，但是由于语音是由人的声道振动而产生的，这种口腔肌肉运动相对于语音频率来说非常缓慢，所以从这一方面看，可以认为语音信号在一个短时间范围内其特性基本保持不变，即相对稳定，因而可以将其看作是一个准稳态过程，这就是语音信号的短时平稳性。

（2）语音信号是功率信号[12]

设连续信号为 $x(t)$，离散信号为 $x(n)$，它们的能量分别定义为：

$$E=\int_{-\infty}^{\infty}\left|x(t)\right|^{2}dt \tag{2-1}$$

$$E=\int_{-\infty}^{\infty}\left|x(n)\right|^{2}dt \tag{2-2}$$

如果 $E<\infty$，那么 $x(t)$ 和 $x(n)$ 就被称为能量有限信号，简称能量信号。

若 $E>\infty$，则称为能量无限信号。对于能量无限信号，一般研究它们的功率。设 $x(t)$ 和 $x(n)$ 的功率分别定义为：

$$P=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left|x(t)\right|^{2}dt \tag{2-3}$$

$$P=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}\left|x(n)\right|^{2} \tag{2-4}$$

若 $P<\infty$，则称 $x(t)$ 和 $x(n)$ 为有限功率信号，简称功率信号。

周期信号、准周期信号以及随机信号，由于它们的时间都是无限的，所以它们总是功率信号，所以语音信号作为一个随时间变化的随机信号来说，它是功率信号而不是能量信号。

（3）清音和浊音[11]

人们根据语音信号的生成机理以及语音信号的特征差异，通常把语音信号分为两大类，一类为清音，一类为浊音。从时域和频域上来看，前者并没有明显的特征，与白噪声非常相似；而后者在时域上表现出非常明显的周期性规律，并且在低频段聚集了其大多数的能量，在频域上的特点是它存在共振峰的特性。

（4）用统计分析特性描述语音信号[11]

统计特性的分析方法是一种非常有效的分析随机信号的方法。由于语音信号是一个随时间变化的非平稳的随机过程,所以可以采用许多种统计分析特性对其进行分析。

2.1.2 噪声特性

在各种各样的信号处理系统中，噪声信号相对于有用信号而言，所造成的都是干扰和破坏作用，但是噪声却是普遍存在的。噪声信号产生于实际的生活环境，存在于实际的生活环境，它的特性随着环境的改变而改变。噪声信号一般分为两类——加性噪声和非加性噪声[11]。

加性噪声一般被分为冲激噪声、周期性噪声、宽带噪声、相同声道情况下其他语音信号的干扰噪声等等[11]。

（1）冲激噪声

放电、打火和爆炸都是产生冲激噪声的原因。此类噪声在时域上表现为波形中突然出现的窄脉冲，这些窄脉冲和冲激函数的表现形式非常相似。

在时域上去除此类噪声时，可先对带噪语音信号的幅度求均值，然后把得到的均值作为判断阈值，超过该阈值的我们认为是冲激噪声，然后将其滤除。

（2）周期性噪声[11]

周期性噪声是由周期性工作的机械设备在使用时发出的一种噪声，例如发动机、电风扇，洗衣机以及周期性运作的机床等等发出的噪声就都属于周期性噪声。另外，一些电气干扰所产生的声音，例如 50Hz 的交流电产生的噪声也属于周期性噪声。此类噪声在频域上表现为很多离散的窄谱，因此可以通过观察信号的功率谱图来发现进而通过滤波方法将其去除。

（3）宽带噪声[11]

宽带噪声的产生原因多种多样，例如量化噪声、随机噪声源产生的各种噪声，呼吸产生的轻微声响等等都属于宽带噪声。此类噪声在时域上的表现是它与语音信号完全重叠在一起，在频域上它的频谱完全遍布于语音信号的频谱当中，因此宽带噪声的消除比较复杂。

（4）语音干扰[11]

当用于传送语音信号的信道中存在着其他的干扰信号，这样就给准备要传送的语音信号造成了干扰，这些干扰信号就称为语音干扰信号。在区分有用语音信号和语音干扰信号的时侯可以利用它们的基音差别。因为在一般情况下，两种语音的基音都不相同，也不会成整数倍的关系，所以可以利用此种差别来提取有用语音信号的基音和各次谐波，再把其恢复即可。

非加性噪声主要是指残响和传输网络的电路噪声等等。这类噪声与背景噪声不同，在时域上它们表现为语音信号与噪声信号的卷积。在实际生活中，加性噪声普遍存在，非加性噪声处理起来比较困难，一般可以通过变换，例如同态变换[13]等，将非加性噪声变换为加性噪声再对其进行处理。

2.2 语音信号处理基础知识

语音信号处理是一门应用数字信号处理技术对语音信号进行处理的学科，它在语音通信、语音合成、语音识别、语音增强等众多的信息处理及应用领域中都是一种很重要的核心处理技术。语音信号处理的目的是把语音信号分析成一些能够表示其本质特征的参数，然后利用这些参数进行相应的应用。本文介绍了一些最基础的语音信号处理知识。

2.2.1 语音信号的短时分析技术

语音信号是一个随时间变化的非平稳信号，所以用来处理平稳信号的一些处理技术对其并不适用。但是因为语音具有短时平稳的特性，所以对语音信号进行处理和分析时，可以将语音信号分为一段一段来分析，其中每一段称为一“帧”，这就是所谓的“语音信号的短时分析”。通常认为语音信号在 10～30ms的时间内相对平稳，因此帧长一般取为 10～30ms。对于整体的语音信号来讲，分析出的是每一帧特征参数所组成的特征参数时间序列[1,11]。

2.2.2 语音信号的预处理

2.2.2.1 预加重

从语音信号的频谱图中可以看出，频率越高的地方，语音信号的成分越小，也就是说语音信号的高频处的频谱比低频处的频谱难求，为此要在语音信号的预处理中进行预加重处理，它的目的是提升高频部分，使语音信号变得平坦，这样就能在低频到高频的整个频带中用同样的信噪比来求频谱。

预加重通常使用具有 6dB/倍频程的提升高频特性的预加重数字滤波器来实现，它一般是一阶有限冲激相应(FIR)滤波器[1,11]：

$$H(z)=1-\mu z^{-1} \tag{2-5}$$

式中 μ 的取值范围是[0.4～1.0]。

2.2.2.2 分帧和加窗

在语音信号的处理中，对进行过预加重处理的信号还要进行加窗和分帧处理。加窗和分帧是实现语音短时分析的手段。

通过加窗处理可以把语音信号分为许多个短时的语音段，每个短时的语音段都被称为一帧。帧和帧之间既可以连续，也可以交叠，但是一般都采用有交叠的分帧方法，这样做的目的是为了使帧与帧之间平滑过渡，保持语音信号的连续性，前一帧和后一帧的交叠部分称为帧移。帧移和帧长的比值一般取为0～0.5之间。图2-1中明确的表示了帧长和帧移的意义[11]。

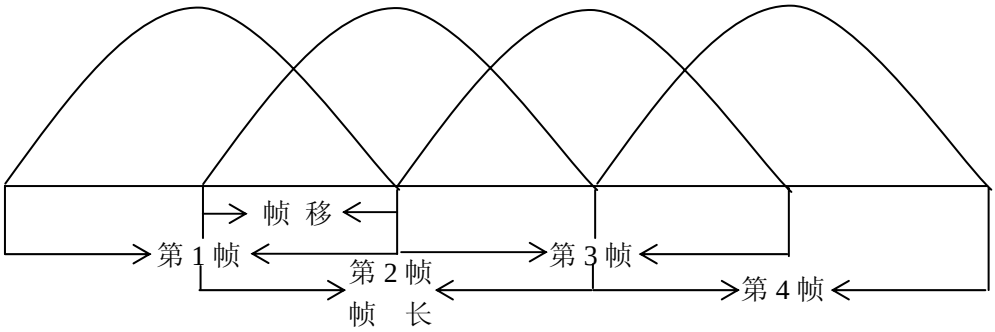


图2-1 帧长和帧移

分帧是用可以移动的窗口进行加权的方法来实现的，窗口的长度是有限长的，设窗函数为 $w(n)$，用窗函数乘以语音信号 $s(n)$，从而形成加窗的语音信号：

$$s_w(n) = s(n) * w(n) \tag{2-6}$$

在语音信号数字处理中最常用的窗函数是矩形窗和汉明窗(Hamming)等，它们的表达式如下[12]：

矩形窗：

$$w(n) = \begin{cases} 1 & n = 0,1,...., \mathrm{N}-1 \\ 0 & n = \text{其它} \end{cases} \tag{2-7}$$

汉明窗：

$$w(n) = \begin{cases} 0.54 - 0.46\cos(\dfrac{2\pi n}{N}) & n = 0,1,..., N-1 \\ 0 & n = \text{其它} \end{cases} \tag{2-8}$$

窗函数的选取对于语音信号的短时参数的影响非常大。为此应该选择合适的窗函数，使语音信号的短时参数能更贴切地反映出语音信号的特征变化。

汉明窗的主瓣宽度为 $\dfrac{8\pi}{N}$，矩形窗的主瓣宽度为 $\dfrac{4\pi}{N}$，由此可知汉明窗的主瓣宽度比矩形窗大一倍，即带宽约增加一倍，同时他的带外衰减也比矩形窗大[12]。虽然矩形窗的谱平滑性能比较好，但是却损失了高频分量，从而造成了波形细节的部分丢失，所以从这一方面来看汉明窗比矩形窗更为适合应用于语音信号的加窗分帧[2]。

2.2.3 语音信号的时域分析

语音信号的时域分析是最早用于分析语音信号的方法，它表达直观，能形象的表现出语音信号的波形特征。相比较而言，时域分析通常用于最基本的参数分析，这种

方法的特点是比较直观、物理意义明确，实现比较方便、运算量少等，所以语音信号的时域分析是一种很重要很常见的语音分析技术。它是对语音信号的时域参数进行分析。最常用的语音时域参数有短时能量，短时过零率，短时自相关函数，短时平均幅度函数等[14]。

2.2.4 语音信号的频域分析

对语音信号进行频域分析，主要是对一些频域的参数进行分析，常用的一些频域参数有频谱、功率谱、倒谱等等，最常用的频域分析方法有傅立叶变换法、线性预测法等。

2.2.4.1 短时频谱

对语音信号进行短时截短，再对截短的语音信号进行短时傅立叶变换，经过变换得到语音信号的短时频谱。求截短后某一帧的语音信号 $s_w(n)$ 的短时频谱为：

$$S_n(e^{j\omega}) = \sum_{m=0}^{N-1} s_n(m) e^{-j\omega m} \tag{2-9}$$

因为此时求得的短时频谱是一段语音信号经过短时傅立叶变换得到，而语音信号被截短之后可认为是平稳信号，因此，通过（2-9）式求得的短时频谱即可认为是一段平稳信号的频谱的近似。

应用短时傅立叶变换来对缓慢时变的频谱来进行分析，是现在最常用的也是最简便的一种分析方法。

2.2.4.2 短时功率谱

可以利用短时自相关函数来求短时功率谱密度，这是因为短时自相关序列的傅立叶变化就是短时功率谱密度，其表达如下：

$$s(w) = \sum_{n=0}^{N-1} r(n) e^{-j\omega n} \tag{2-10}$$

其中 $r(n)$ 为短时自相关序列[15]。

第 3 章 语音端点检测算法的分析与研究

从一段语音信号中把语音的起始点和结束点寻找出来，进而划分出语音段和非语音段，这就是所谓的语音端点检测。随着语音处理技术的提高，越来越多的语音端点检测方法开始被人们利用，各种算法都有不同的特点和应用范畴，常用的检测算法包括短时能量检测法、短时幅度检测法、短时过零率检测法、短时自相关检测、语音熵检测算法等。本章将对各种算法进行分析，进而提出一种鲁棒性好的语音端点检测算法。

3.1 短时能量检测法

短时能量分析的依据是在一段包含噪声的语音信号中，语音段的能量为语音信号的能量与噪声信号的能量的叠加，因此语音段的能量就会比非语音段的能量大，因此利用此点就可进行语音端点检测。短时能量检测算法又分为短时能量检测法和短时平均幅度检测法两种检测方法[16]。

3.3.1 短时能量检测

语音信号经过加窗分帧处理后，得到加窗语音信号为公式（3-1）所示：

$$s_w(n) = s(n) * w(n) \tag{3-1}$$

设第 n 帧语音信号的短时能量为 E_n，它等于该帧语音信号取样值的平方和，用公式表示为：

$$E_n = \sum_{m=0}^{N-1} s_w{}^2(m) \tag{3-2}$$

因为短时能量 E_n 的计算是建立在对取样值平方后再进行求和的基础之上的，所以当语音信号的幅度发生变化时，E_n 会发生特别明显的变化，尤其是对语音信号的高电平的变化十分敏感。由于该种检测算法中有平方运算，从而就使得振幅不相等的相邻采样值之间的幅度差别增大，这样在选择窗函数的时候，就必须选择窗宽比较宽的窗，因为只有选择宽度比较宽的窗函数才能对这种起伏变化很大的幅度具有好的平滑作用，但是如果窗宽过宽的话，语音信号的能量时变的特性就不容易反映出来，从而短

时能量检测方法在对于窗函数的窗宽如何选择来说比较困难[17]。

3.3.2 短时幅度检测

鉴于短时能量对高电平十分敏感，因此可采用另外一个时域参数——短时幅度函数，设为M_n，该函数也是用来度量语音信号幅度值变化的，把它表示为：

$$M_n = \sum_{m=0}^{N-1} \left| s_w(m) \right| \tag{3-3}$$

它也代表第n帧语音信号的能量的大小，由M_n的表达式可以看出，取样值的大小不会因为平方而造成较大的差异，这是它与E_n的最大不同之处，这就避免了短时能量E_n对高电平敏感的缺陷。

短时能量和短时平均幅度的主要优势在于[17,18]：

（1）检测方法简单，容易实现；

（2）可以对清音段和浊音段进行很好的检测，因为根据语音信号的特点，浊音的能量比清音的能量要大很多；

（3）能用其进行无声段和有声段的分界，能进行声母和韵母的分界；连字（连字指的是字与字之间不存在间隙）的分界等。

所以该种方法可以作为一种语音信号的特征参数应用于语音检测中，而且它也是最传统的最简单的一种检测方法。但是它也存在一定的缺陷：

（1）不能很好的排除单音频信号；

（2）该种方法在背景噪声增加或者变化的情况下，它的漏检和误检率都比较高，在低信噪比的情况下不能有效的进行语音端点检测。

3.2 短时过零率检测

语音信号的过零分析就是指语音信号通过零值，它是语音时域分析中非常简单的一种分析方法，对于连续的语音信号来说，研究其过零分析就是观察其时域波形通过时间轴的情况。对于离散后的语音信号，如果相邻的取样值改变符号则认为是过零。由此可以计算过零的次数，过零次数就是样本改变符号的次数，即为过零率。

设语音的短时过零率为Z_n，定义为：

$$Z_n = \frac{1}{2}\sum_{m=0}^{N-1}\left|\mathrm{sgn}\left[s_w(m)\right]-\mathrm{sgn}\left[s_w(m-1)\right]\right| \quad (3\text{-}4)$$

其中$\mathrm{sgn}\left[x(n)\right]$为符号函数，表示为：

$$\mathrm{sgn}\left[x(n)\right]=\begin{cases} x(n)=1 & x(n)\geq 0 \\ x(n)=-1 & x(n)<0 \end{cases} \quad (3\text{-}5)$$

由此表达式可以看出，如果相邻两个取样值为同号，那么可求得的：

$$\mathrm{sgn}\left[x(n)\right]=0 \quad (3\text{-}6)$$

如果相邻两个取样值为异号，那么求得的值为：

$$\mathrm{sgn}\left[x(n)\right]=1 \quad (3\text{-}7)$$

短时过零率的优点是可以从掺杂着噪声的语音信号中找出有效的信号，可以准确的判断有声段和无声段的起始点和结束点，尤其是对于语音识别系统中的孤立词的检测有着更为重要的作用。跟短时能量检测方法相比起来，短时能量方法在背景噪声较小的情况下比较准确，而在背景噪声稍微增大的环境下，短时过零率的检测方法更为适用。

虽然短时过零率的方法计算简单，易于实现，但是它同短时能量检测法一样也不能有效的排除单频信号，而且它对清音和浊音的区别也不是特别准确，其检测结果的好坏也容易受到背景噪声的影响，这些都是短时过零率检测算法的缺点[4,17]。

3.3 短时自相关检测

用短时自相关函数作为分析方法来进行语音检测的依据是噪声信号的相关性很小，求得的自相关函数非常小，但是语音信号的相关性很强[15]，所以可以利用自相关性来区分噪声段和语音段，从而进行语音端点检测。

虽然可以利用短时自相关函数来进行语音端点检测，但是跟短时过零率以及短时能量的检测方法比较起来，它的计算非常复杂，使得运算量加大，在实际应用中并不多见。

3.4 倒谱距离检测法

倒谱特征是语音信号频域参数中的一个，它能很好的反映出语音信号的特点，因此应用此种特征来进行语音端点检测也是可行的。倒谱距离检测法的原理是因为对数

谱的均方距离可以表示两个信号谱的区别，因此它可以作为一个特征参数来判断该信号是语音信号还是噪声信号[19,20]。

设语音信号为 $s(n)$，它的倒谱变换设为 $c(n)$。

语音信号的倒谱指的是信号能量谱密度函数 $s(\omega)$ 的对数的傅立叶变换，定义表示如下[21]：

$$l\log s(\omega)=\sum_{n=-\infty}^{\infty}c(n)e^{-jn\omega} \tag{3-8}$$

式中，$c(n)$ 为倒谱系数，当为实数的时候有：

$$c(n)=c(-n) \tag{3-9}$$

$c(0)$ 的定义为：

$$c(0)=\int_{-\pi}^{\pi}\log s(\omega)\frac{d\omega}{2\pi} \tag{3-10}$$

对于一对能量谱密度函数 $s(\omega)$ 和 $s'(\omega)$，根据 Parseval 定理，对数谱的均方距离可以用倒谱距离来表示，其公式为：

$$d_{cep}{}^{2}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\log s(\omega)-l\log s'(\omega)\right|^{2}d\omega=\sum_{n=-\infty}^{\infty}(c(n)-c'(n))^{2} \tag{3-11}$$

式中 $c(n)$ 和 $c'(n)$ 分别是对应于谱密度函数 $s(\omega)$ 和 $s'(\omega)$ 的倒谱系数。

用测量倒谱距离的方法来判断每一帧信号是语音信号还是噪声信号，根据每一帧信号与噪声的倒谱距离的轨迹就可以进行语音端点检测。

此种检测方法的缺陷是计算复杂，运算量大，应用于语音端点检测不好解决实时问题[22]。

第 4 章 基于语音熵的端点检测方法的研究与仿真

传统的语音端点检测算法，如基于短时能量以及短时过零率的检测方法，虽然计算简便，但是在低信噪比的情况下，该算法的检测效果就会很差，基于模式识别的方法准确性较好，但是相对来说计算量大，运算复杂，很难应用于语音信号处理系统当中去[23]。为了解决语音信号能量小易被淹没以及避免大量运算，本章介绍一种基于语音熵的语音端点检测算法。

4.1 熵的概念及定义

信息熵的概念由Shannon首次提出，在信息论中，用熵函数表征信源输出的平均信息量。因为平均自信息量的表达式与物理学中的热熵的表达式很类似，所以将其命名为熵，定为信息熵，又称仙农熵[24]。

假设信源发出有限个符号，它们组成的输出序列前后符号之间相互统计独立，N 个符号出现的概率分别为 $P_1, P_2, P_3 ..... P_N$，则信息熵定义为[24]：

$$H(X) = E[\log \frac{1}{P_i}] = -\sum_{i=1}^{N} P_i \log P_i \qquad (4\text{-}1)$$

其中

$$\sum_{i=1}^{N} P_i = 1 \qquad (4\text{-}2)$$

以上是以信源的消息是离散的为例来定义信息熵的，但是现在要研究的语音信号是连续的消息，为了计算语音信号的信息熵，可以先对语音信号进行离散处理，即对语音信号进行采样，然后对离散后的语音信号求其信息熵。

4.2 基于熵的语音端点检测算法

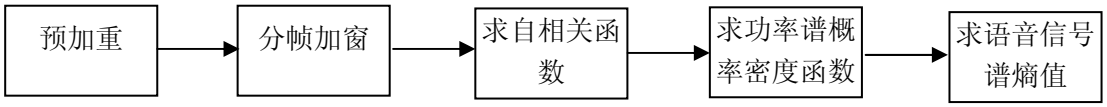
信源熵仅与符号的概率分布有关，是概率分布的函数，当信源 N 个符号的分布概率相等时，其熵值最大。这说明信源符号的概率分布越平坦，则熵值越大，其包含的平均信息量越大[17]。

Shen在实验室中发现语音信号的熵和噪声信号的熵存在较大的差异，于是首次将熵引入语音信号的检测中。对于一段频带很宽的语音信号来说，由于语音段相对于背景噪

声而言，它的能量主要集中在某几个频段，起伏突变比较大，所以熵值小，而噪声信号在整个频带内分布相对比较平坦（尤其是白噪声信号），所以其熵值比较大，因此可以利用这种差异来区分语音段与噪声段[25]。对于频带宽度受限的情况而言，例如语音信号的频率主要集中在300Hz~3400Hz，所以在此范围内，则认为语音信号的随机事件比较多，平均信息量大，熵值大，在背景噪声在此频率范围受限的情况下，则熵值较小。本文中所有的仿真结果都是在频率受限（限制在300Hz~3400Hz的范围内）的情况下经过实验所获得。

Shen提出的是利用语音信号直接变换所得的频率谱进行语音信号的端点检测，然而，由于语音信号是功率信号而不是能量信号，一般对语音信号进行频域分析都是分析的语音信号的功率谱密度，所以基于上述原因，本课题中用语音信号的短时功率谱来取代短时频谱来构造语音信息熵，这样就能更好的进行语音端点检测，能更好的对语音段和噪声段进行区分。

该算法最简单的实现框图4-1所示：



4-1 实现框图

其具体步骤如下：

设用 $s(n)$ 来表示一段语音信号。

（1）首先对语音信号 $s(n)$ 进行预加重；

（2）然后通过加入窗函数，对其分帧、加窗，设窗函数为 $w(n)$，加窗之后语音信号的表达式为：

$$s_w(n) = s(n) * w(n)\ ;\quad n = 0,1,...,N-1 \tag{4-3}$$

（3）求每一帧的短时自相关函数；

$$r(n) = \sum_{i=0}^{N-n-1} s_w(i)s_w(i+n) \qquad n = 0,1,...,N-1 \tag{4-4}$$

（4）对上述自相关函数进行快速傅立叶FFT变换，得到语音帧的短时功率谱密度，如公式（4-5）所示：

$$s_w(i,m)=\sum_{n=0}^{N-1}r(n)e^{-j\frac{2\pi kn}{N}} \qquad i=0,1,...,N-1 \tag{4-5}$$

其中 $s_w(i,m)$（$i=0,1,2,...,N-1$）表示第m帧第 i 频率分量的功率谱密度或者称为第m帧第 i 频率分量的功率谱幅度；

（5）由此定义和计算在这一帧中的每个频率分量的归一化功率谱概率密度函数（pdf）：

$$P(i,m)=\frac{s_w(i,m)}{\sum_{i=0}^{N-1}s_w(i,m)} \qquad i=0,1,...,N-1 \tag{4-6}$$

上式是每一个频率分量的功率谱能量占整个这一帧的功率谱能量的概率。对于实信号来说，其N点FFT变换是关于N/2+1点对称的，即功率谱密度有对称性，所以只取一帧中一半的点来构造概率密度函数 $P(i,m)$ 就可以，所以 $P(i,m)$ 又可写成：

$$P(i,m)=\frac{s_w(i,m)}{\sum_{i=0}^{N/2-1}s_w(i,m)} \tag{4-7}$$

（6）由于语音信号的频率主要集中在300~3400范围内，因此可以加入对频段范围的限制条件，如下所示：

$$s_w(i,m)=0 \qquad f_i<300Hz,\ f_i>3400Hz \tag{4-8}$$

f_i 代表 i 点的频率。加入此约束条件后，使得语音段和非语音段的区分能力增强。

（7）计算得出此帧语音信号的短时信息熵。

$$H(m)=\sum_{i=0}^{N/2-1}P(i,m)*\log[1/P(i,m)]=-\sum_{i=0}^{N/2-1}P(i,m)*\log P(i,m) \tag{4-9}$$

由（4-9）式可以看出，语音谱熵具有以下特点：

（1）语音谱熵只与语音信号的随机性有关，而与语音信号的幅度无关，理论上认为只要语音信号的分布不发生变化，那么语音谱熵不会受到语音幅度的影响。另外，由于每个频率分量在求其概率密度函数的时侯都经过了归一化处理，所以从这一方面也证

明了语音信号的谱熵只会与语音分布有关，而不会与幅度大小有关[26]。图4-2是对此进行的仿真图。从图上可以看出，第一个语音“yi”的幅度比第二个语音“yi”要大很多，但是它们的谱熵却没有很大的差别。

（2）在一定程度上，语音谱熵与噪声谱熵的区别十分明显，应用于语音端点检测具有一定的鲁棒性[27]。但是，当信噪比降低的时候，情况会发生变化。图4-3为同一段语音“yi”在不加入噪声，以及分别加入SNR为10dB、20dB噪声下的谱熵曲线，可见当信噪比降低时，语音信号的谱熵值的形状大体保持不变，这说明谱熵是一个比较稳健性的特征参数，但是随着信噪比的降低，其谱熵值也降低，且语音信号的谱熵值与噪声信号的谱熵值差别变小，因此在恶劣的噪声环境下，利用谱熵进行语音端点检测变的比较困难[28,29]，需要对语音信号先进行滤波降噪处理，以提高其信噪比。

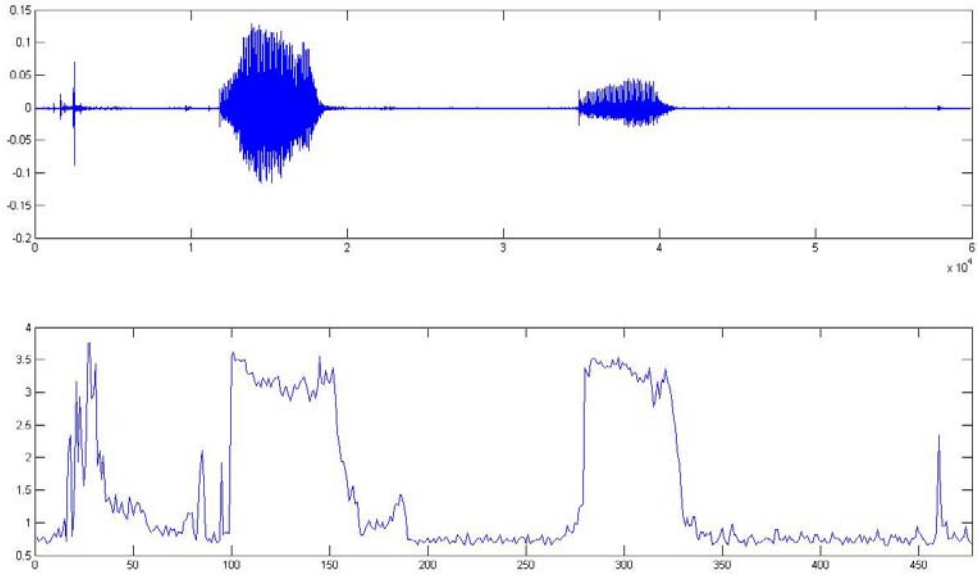


图4-2　语音“yi”的幅度大小对功率谱熵的影响

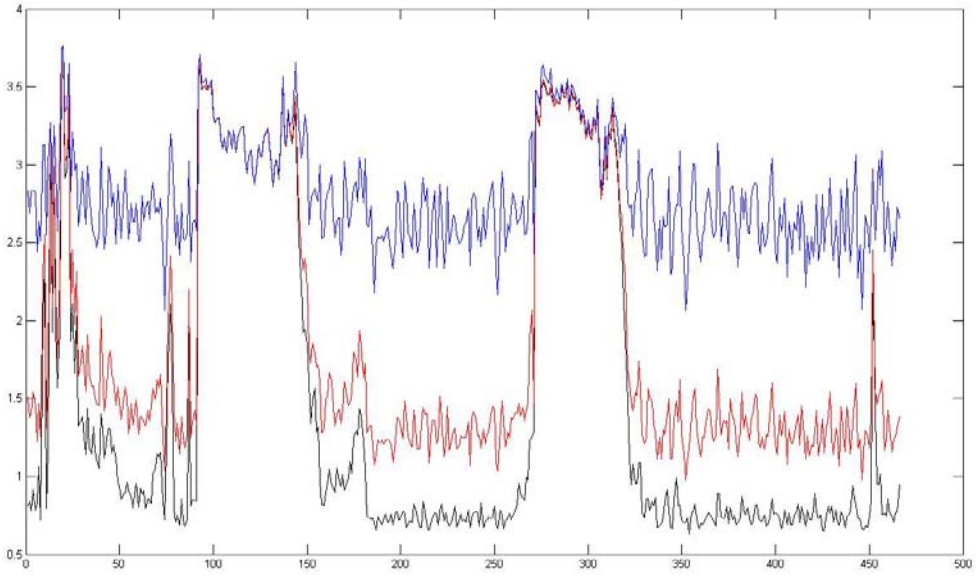


图4-3 语音“yi”在不同噪声下的谱熵值

4.3 自适应子带谱熵

4.3.1 子带谱熵

从 4.2 节中的语音信号的谱熵计算公式可以看出，由功率谱密度函数构造的谱熵只依赖于功率谱能量的变化，而不是依赖谱的能量，换句话说，也就是功率谱熵只与谱的能量的变化有关而与谱的能量无关，所以可以看出，由功率谱密度函数构造的谱熵在不同的噪声环境下有一定的鲁棒性。但是图 4-3 中表明，在背景噪声存在的情况下，语音信号容易受到噪声信号的干扰，导致谱熵不能对语音信号和噪声信号进行很好的区分。

不同噪声的频率集中在不同的频率子带上，因此文献[30]提出了应用多子带技术对带噪语音信号进行分析的方法。多子带分析技术不仅能消除被噪声污染的有害子带，而且能更好的捕捉语音信号的本质特征[30]。所以Bing-Fei Wu，Kun-Ching Wang等就把多子带分析技术应用到计算谱熵的过程中，进而提出了一种基于子带谱熵的语音端点检测算法[31]。

基于子带功率谱概率密度函数构建的谱熵与基于单个功率点概率密度函数构建的谱熵的不同之处在于，后者是求每一帧信号中的每个功率点的概率密度函数，然后构

建谱熵，而前者是先把一帧信号分成若干个子带，求的是每一个子带的概率密度函数，进而再进行谱熵的计算，这样就消除了单个功率点上的幅度易受噪声影响的问题[16]。

设语音信号的第 m 帧的第 l 子带的功率谱能量为 $E(l,m)$，将第 m 帧语音信号分成 N_a 段，即 N_a 个子带，且 $N_a=\frac{N}{b}$，其中 N 为此帧的长度，b 为划分的子带的长度。将 $E(l,m)$ 定义如下：

$$E(l,m)=\sum_{k=1+(l-1)*\frac{b}{2}}^{k=1+(l-1)*\frac{b}{2}+(\frac{b}{2}-1)} s_w(k,m) \qquad 1\le l\le N_a \tag{4-10}$$

式中取 $\frac{b}{2}$ 的原因也是由于 s_w 的对称性。

令 $N=256$，$b=8$，则：

$$N_a=\frac{N}{8}=32 \tag{4-11}$$

那么 $E(l,m)$ 为：

$$E(l,m)=\sum_{k=1+(l-1)*4}^{k=1+(l-1)*4+3} s_w(k,m) \qquad 1\le l\le 32 \tag{4-12}$$

基于上述子带功率谱能量，定义子带功率谱概率为：

$$P(l,m)=\frac{E(l,m)}{\sum_{k=1}^{N_a}E(k,m)} \qquad 1\le l\le N_a \tag{4-13}$$

子带功率谱熵为：

$$H(m)=\sum_{l=1}^{N_a}P(l,m)*\log[1/P(l,m)]=-\sum_{l=1}^{N_a}P(l,m)*\log P(l,m) \tag{4-14}$$

图 4-2 表明的是基于单个功率点的概率密度函数构建的谱熵值的仿真图，图 4-4 是基于子带功率谱概率密度函数构建的谱熵值的仿真图，可见二者有明显的区别。

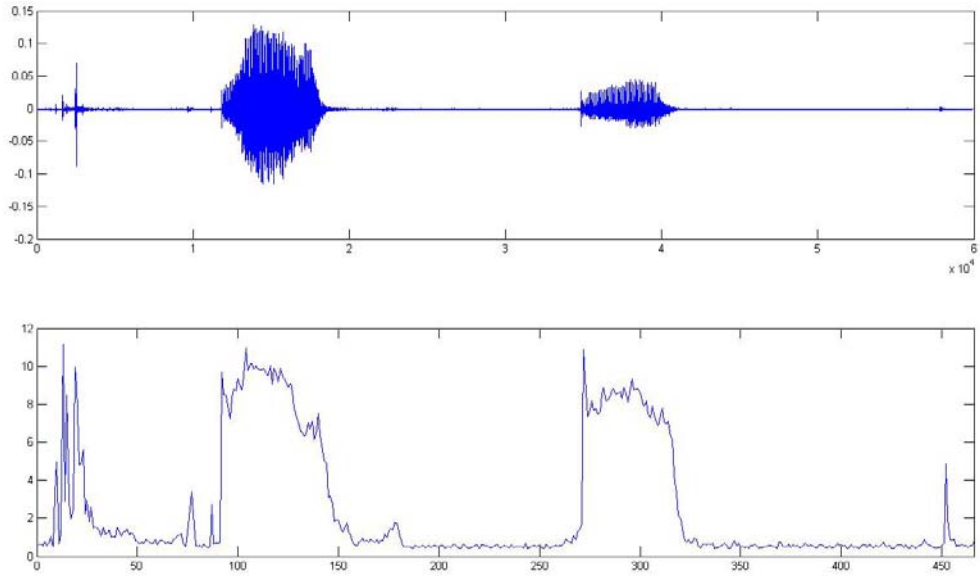


图 4-4 基于子带功率谱构建的谱熵值

4.3.2 子带自适应选择

在构成子带功率谱熵时，并不是每一帧都分割成相同的子带数目，因为输入的语音信号是随机的，如果每一帧的子带数目相同的话，那么这种子带谱熵的优越性并不能很好的体现。为了确保语音端点检测算法能达到实时处理以及在恶劣噪声环境下还能保持其稳健性，所以提出了一种子带自适应选择的方法[31]，让每一帧的语音信号划分的子带数目都不尽相同，这样更能符合语音信号的本质。

噪声能量大的子带对语音信号的干扰比较严重，这种子带被称为有害子带。有害子带的数目和有用子带的数目是由背景噪声的水平所决定的。子带自适应选择的原理是基于文献[32]中最小化子带功率谱能量而提出的，最小化子带功率谱能量能用于背景噪声水平的估计，但是此参数对噪声的变化特别明显，为了解决此种问题，提出了归一化最小子带功率谱能量，该参数能精确的对连续语音信号进行噪声估计[32]，如果该参数取值越小说明噪声越小，该参数越大说明噪声越大。

第m帧语音信号的归一化最小子带功率谱能量定义为：

$$NMinB(m) = -\log \frac{\min(E(l,m))}{\sum_{l=1}^{N_a} E(l,m)} \tag{4-15}$$

设 N_u 代表每帧语音信号的有用子带数目，在文献[31]中， N_u 与归一化最小子带功率谱能量的关系表达式为：

$$N_u=\begin{cases}30 & NMinB<5\\ 4+\dfrac{NMinB-5}{25-5}*(30-4) & 5<NMinB<25\\ 4 & NMinB>25\end{cases} \tag{4-16}$$

则第m帧语音信号的子带自适应选择功率谱密度函数为：

$$P(l,m)=\frac{E(l,m)}{\sum_{l=0}^{Nu}E(l,m)} \tag{4-17}$$

从而得到的基于子带自适应选择的功率谱熵计算公式为：

$$Hb(m)=\sum_{l=1}^{Nu}P(l,m)*\log[1/P(l,m)] \tag{4-18}$$

图4-5、图4-6、图4-7以及图4-8中给出的是一段噪声变化的语音信号，分别对基于相同子带功率谱构建的谱熵以及基于子带自适应选择法构建的谱熵进行的仿真，可见在加入信噪比的噪声信号情况下，基于子带自适应选择法构建的谱熵能更好的表征出语音信号的谱熵值。

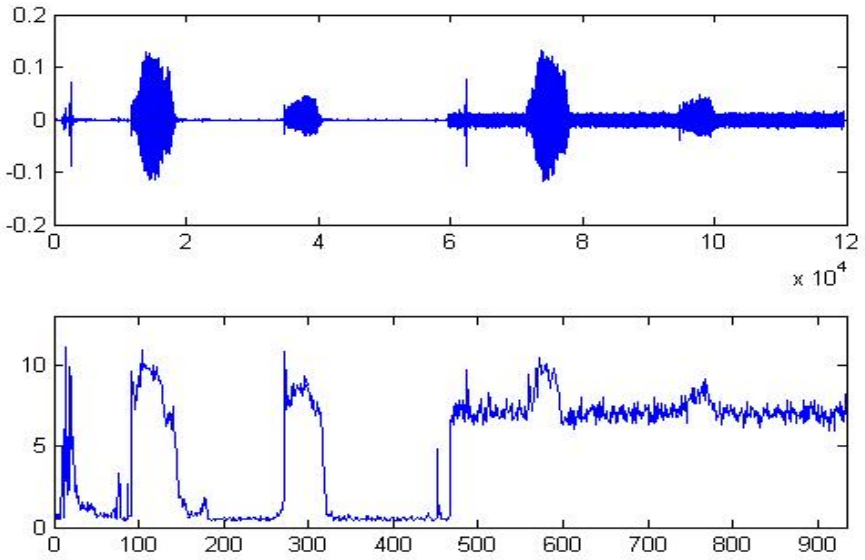


图4-5 加入20dB噪声基于相同子带功率谱构建的谱熵

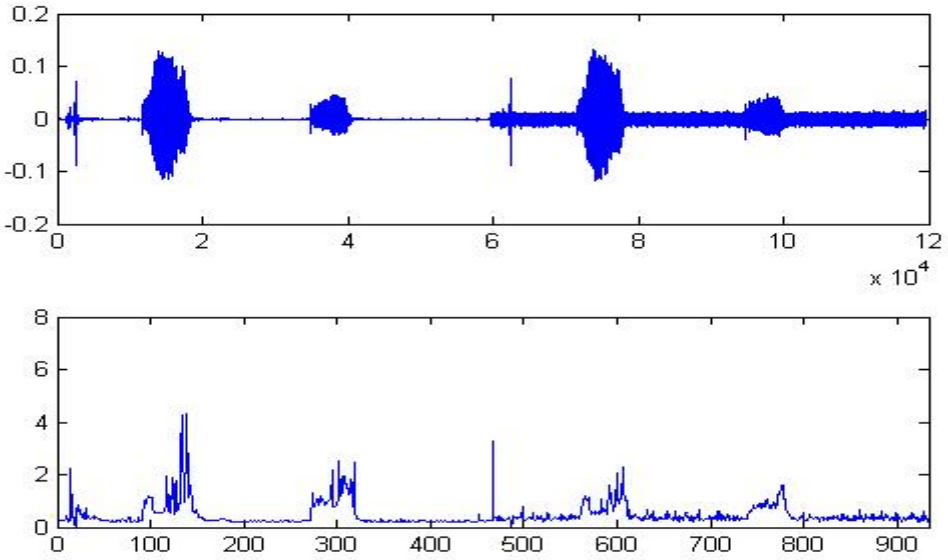


图4-6 加入20dB噪声基于文献[31]中的子带自适应选择功率谱构建的谱熵

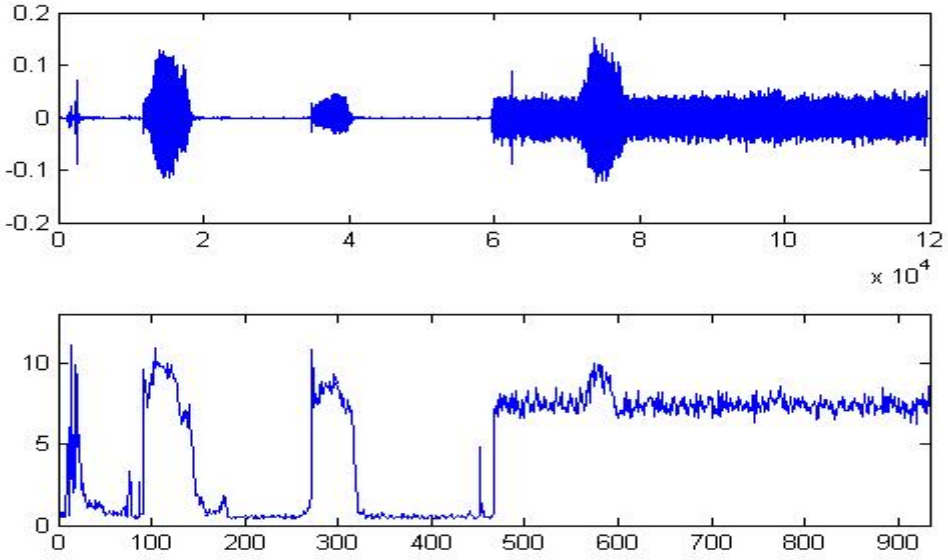


图4-7 加入10dB噪声基于相同子带功率谱构建的谱熵

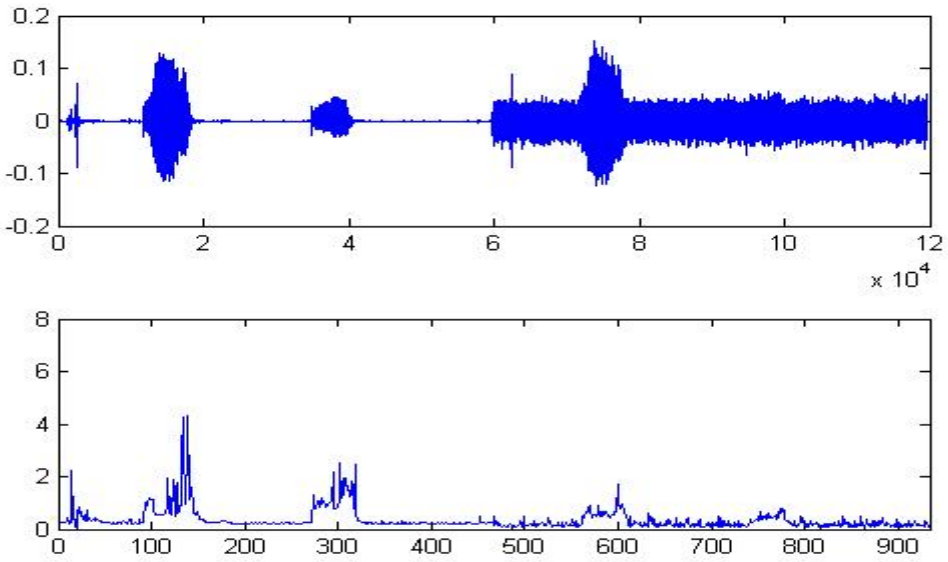


图4-8 加入10dB噪声基于文献[31]中的子带自适应选择功率谱构建的谱熵

文献[33]对 N_u 的公式进行了改进，如公式（4-19）所示：

$$N_u=\begin{cases}32 & NMinB<5\\ 36.5+\dfrac{NMinB}{25-5}*(6-32) & 5<NMinB<25\\ 6 & NMinB>25\end{cases} \tag{4-19}$$

由于在本论文中也对 N_u 的公式进行了改进，如公式（4-20）所示：

$$N_u=\begin{cases}25 & NMinB<5\\ 10+\dfrac{NMinB-5}{25-5}*(25-10) & 5<NMinB<25\\ 10 & NMinB>25\end{cases} \tag{4-20}$$

由图4-6以及图4-8的仿真可以看出，文献[31]中提出的子带自适应选择公式计算得出的谱熵较小，所以随着噪声的增大，谱熵特征容易被噪声淹没。而文献[33]中提出的改进方法经过图4-10与图4-13仿真，并与图4-5和图4-7进行比较可知，它改进后的效果类似于基于相同子带功率谱熵的效果，其原因是它划分的子带数目比较过于接近相同子带功率谱熵中选用的子带数目。经过对本文中改进的子带自适应选择公式进行的仿真可知，语音信号的谱熵与噪声信号的谱熵在本文改进的方法下具有很明显的差异，且不易被噪声所淹没。仿真效果如下所示：

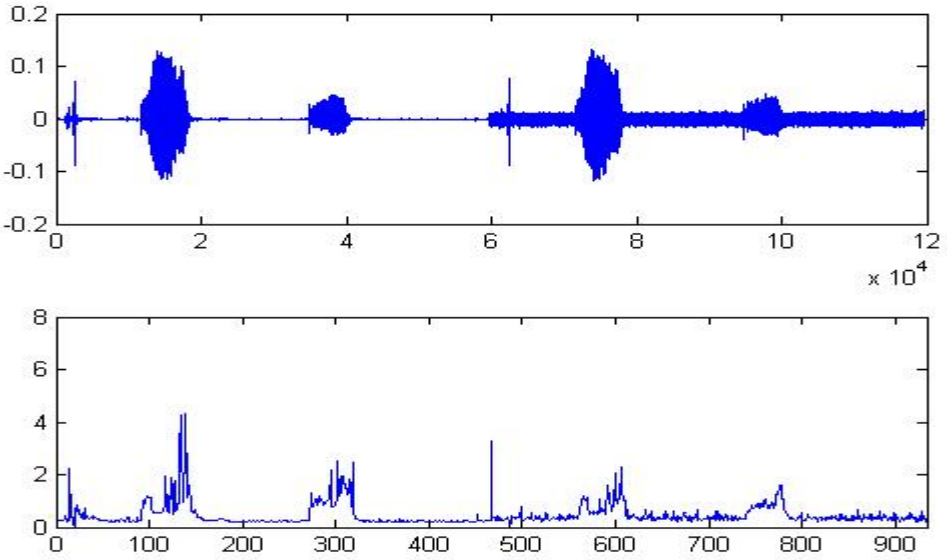


图4-9 加入20dB噪声基于文献[31]中的子带自适应选择功率谱构建的谱熵

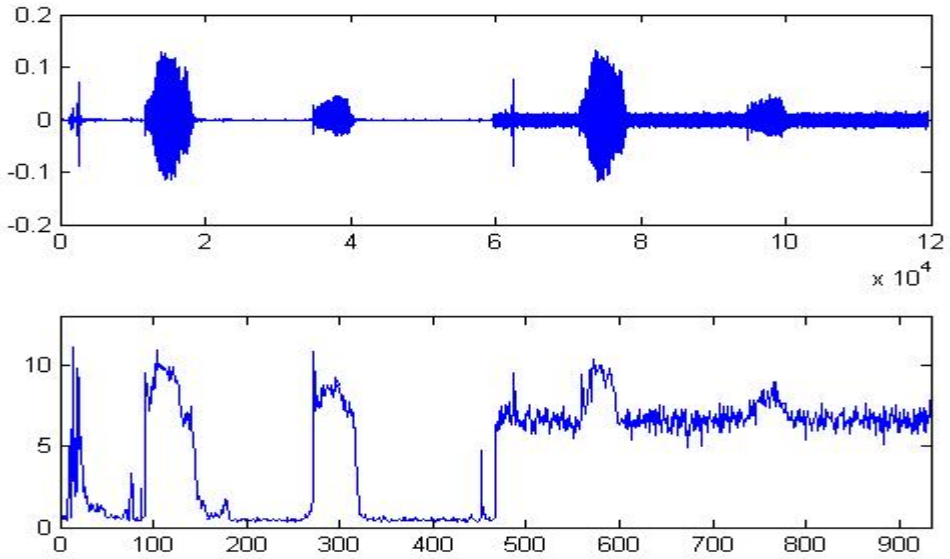


图4-10 加入20dB噪声基于文献[33]中的子带自适应选择功率谱构建的谱熵

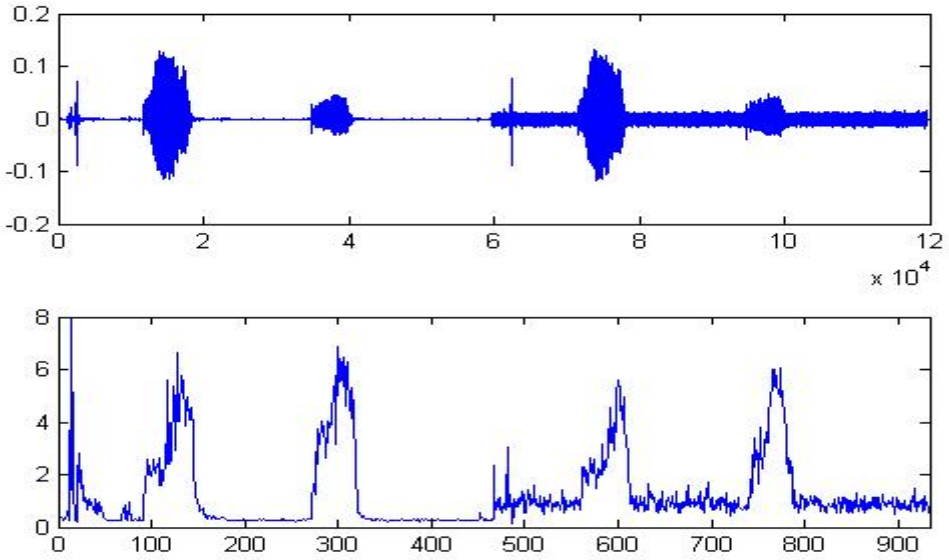


图4-11 加入20dB噪声基于本论文中改进的子带自适应选择功率谱构建的谱熵

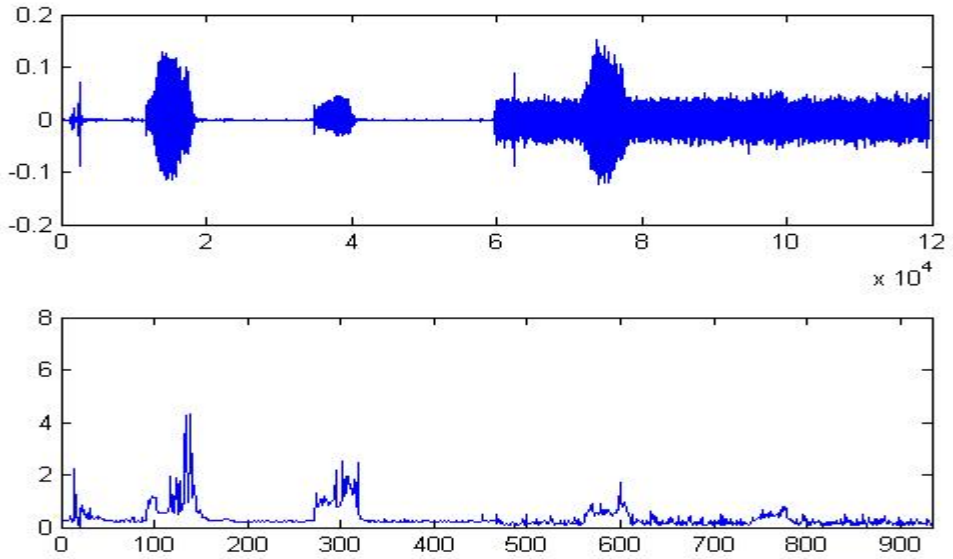


图4-12 加入10dB噪声基于文献[31]中的子带自适应选择功率谱构建的谱熵

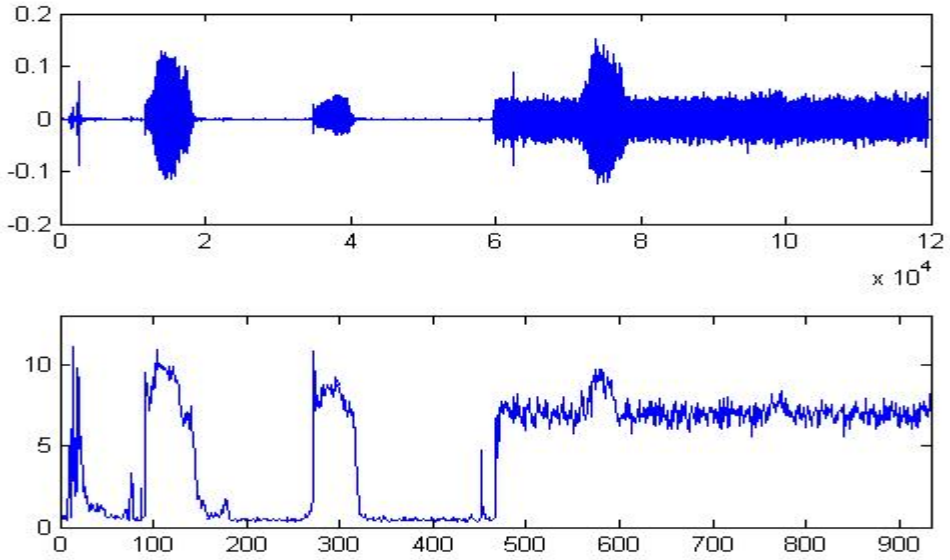


图4-13 加入10dB噪声基于文献[33]中的子带自适应选择功率谱构建的谱熵

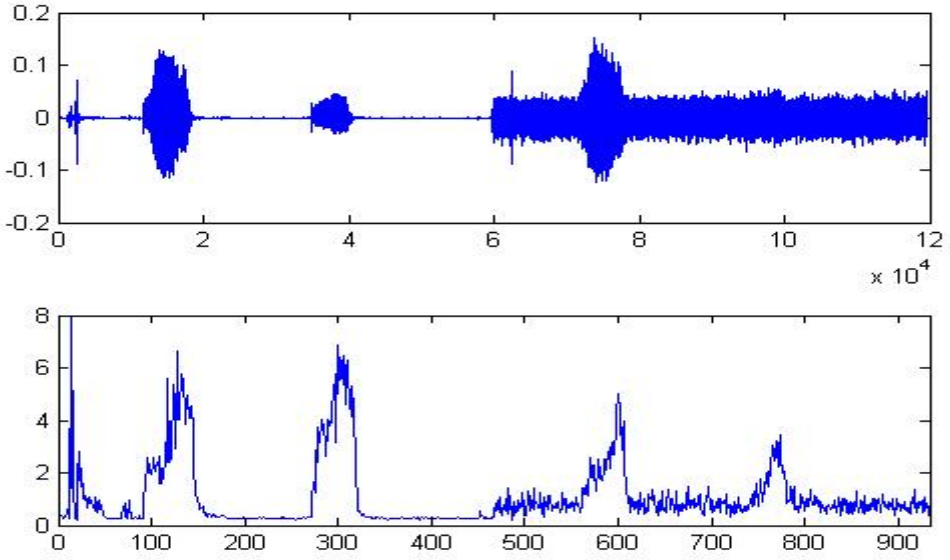


图4-14 加入10dB噪声基于本论文中改进的子带自适应选择功率谱构建的谱熵

第 5 章 自适应滤波技术

5.1 自适应滤波的基础理论

自适应滤波理论在 20 世纪 60 年代提出，是近四十年来发展起来的一种最佳滤波算法，它的提出对信息处理以及信号处理的发展有着十分重要的作用。自适应滤波器之所以得此称谓，是因为它自身可以自动调整内部结构以及参数，从而达到改善其对信号处理能力的目的[34]。

自适应滤波器的优点是可以在对外界环境无先验知识的条件下，对各种信号进行滤波处理。换句话说，也就是自适应滤波器能在输入信号与噪声的统计特性未知的情况下，自动估计出所需的统计特性，并且可以自动调整滤波参数，达到最佳的滤波效果[35]。

5.1.1 自适应滤波器的发展现状

经过多年的研究，自适应滤波理论在很大程度上得到了发展，它主要发展成为四个理论分支：维纳（Wiener）滤波器理论、卡尔曼（Kalman）滤波器理论、最小二乘（Least Square，LS）估计的方法以及神经网络的方法[34]。

维纳（Wiener）滤波器理论和卡尔曼（Kalman）滤波器理论都是基于最小均方误差推导得到的，而且卡尔曼滤波器是对维纳滤波器的发展，它们都是对最小均方误差的最佳估计，维纳滤波应用于平稳信号的最优预测和滤波，卡尔曼在信号是非平稳的情况下也能取得很好的效果。

基于最小均方误差准则的，最小二乘估计算法不同于维纳滤波算法和卡尔曼滤波算法，它是以最小误差平方和为优化目标。递归最小二乘（Recursive Least Square，RLS）算法是它的代表算法。

自适应滤波器可以分为线性自适应滤波器和非线性自适应滤波器。非线性自适应滤波器的典型代表就是基于神经网络的自适应滤波器，神经网络是一种模拟生物神经模型信号处理能力的计算结构。生物学中，神经网络是一个由大量神经元相互联结形成的网络系统，此处所说的基于神经网络的自适应滤波器实质上是一个高度的非线性的动力学网络系统。这个系统具有很多优势，例如：自适应性、自训练、自学习、自

组织能力等，另外它还有容错性、并行性、稳健性等特点，因而它可以实现许多传统的信号和信息处理技术所不能完成的事情。

非线性自适应滤波器有很强的信号处理能力。但是非线性自适应滤波器的计算较复杂，实际用得最多的仍然是线性自适应滤波器[35]。

5.1.2 自适应滤波器的基本结构

图 5-1 为自适应滤波器的一般结构。 从图中可以看出，它主要由两部分构成——可编程滤波器和自适应算法。图中可编程滤波器是一个可以通过调整自身参数改变滤波状态的滤波器，自适应算法的作用是设立调整参数的原则。其中，k 代表迭代次数，x(k)表示输入信号，y(k)为自适应滤波器输出信号，d(k)定义了期望响应信号（简称期望信号）。

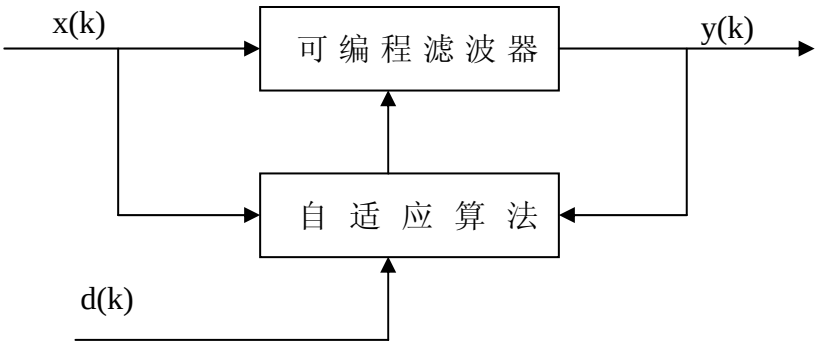


图 5-1 自适应滤波器的基本结构

根据滤波器的输入信号 x(k)，输出信号 y(k)和期望信号 d(k)构造一个目标函数，这个目标函数可以看成是某个误差信号的直接函数，令误差函数为 e(k)，其表达式为：

$$e(k) = d(k) - y(k) \qquad (5\text{-}1)$$

通过将目标函数最小化的过程，用某种方式调整滤波器的参数或者结构，最终使自适应滤波器的目标函数最小化，这样就能够让自适应滤波器的输出信号与期望信号最好的匹配，达到最佳滤波效果，这就是自适应滤波器的基本思想[36]。

总之，所谓自适应滤波，就是利用前一时刻已获得的滤波器参数等结果，自动调节此时刻的滤波器参数，以适应信号和噪声未知或随时间变化的统计特性，从而实现最优滤波。此滤波方法既适用于平稳随机信号也适用于非平稳随机信号。

5.2 自适应滤波器

可编程滤波器以及自适应滤波器算法是组成自适应滤波器必不可少的两部分。本节将分别对这两部分进行介绍。

5.2.1 自适应滤波器结构

可编程滤波器结构即为自适应滤波器的结构。主要包括无限冲激响应（Infinite-duration Impulse Response，IIR）滤波器和有限冲激响应（Finite-duration Impulse Response，FIR）滤波器这两种类型。滤波器结构的选择对算法的处理起着重要的影响。

IIR 型滤波器既包括正向通路又包括反馈通路，也就是说滤波器的输出在反馈支路存在时就会回到滤波器的输入端，并作为输入进入滤波器，因此 IIR 型滤波器的传输函数既有零点又存在极点，因此可以用不高的阶数就可以实现具有陡峭边沿的通带特性，但是由于反馈的存在，使得滤波器的稳定性降低，导致滤波器震荡，另外 IIR 型滤波器还存在收敛速度慢且相位特性难于控制的缺点。

FIR 滤波器是全零点滤波器，也就是说在 FIR 滤波器结构中只存在正向通路，滤波器的输出由当前及过去输入信号的线性组合而构成，这样就使得滤波器的脉冲响应是有限域的，因而它始终是稳定的，且能实现线性的相位特性。由此可以得出，自适应滤波器的结构应采用 FIR 型滤波器。在现实应用中一般采用横向 FIR 滤波器结构，如图 5-2 所示：

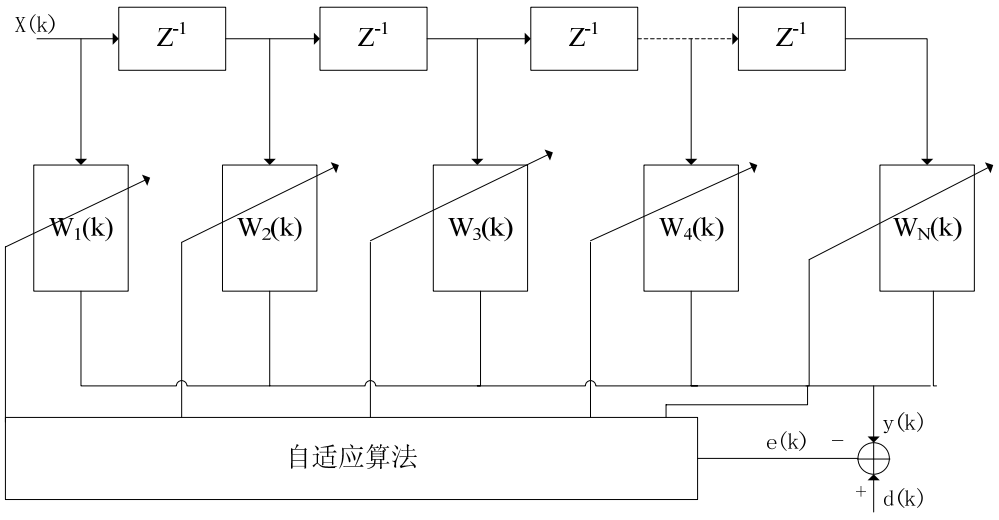


图 5-2 横向 FIR 滤波器结构

（1）图中$W_1(k)$、$W_2(k)$、$W_3(k)$.....$W_N(k)$分别代表各个滤波器所在时刻的权系数。

（2）调节权系数的过程，针对此 FIR 滤波器来说，对权系数进行调节即是使目标函数最小化的方法。

令权系数$W(k)$为：

$$W(k)=[w_1(k),w_2(k),w_3(k),w_4(k),...,w_N(k)]^T \tag{5-2}$$

令输入信号$X(k)$为：

$$X(k)=[x(k),x(k-1),x(k-2),x(k-3),...,x(k-N+1)]^T \tag{5-3}$$

则输出信号$y(k)$为：

$$y(k)=X(k)^TW(k)=W(k)^TX(k) \tag{5-4}$$

图中$e(k)$为误差序列，其表示如下：

$$e(k)=d(k)-y(k) \tag{5-5}$$

把公式（5-4）带入公式（5-5）得：

$$e(k)=d(k)-y(k)=d(k)-X(k)^TW(k) \tag{5-6}$$

其中$d(k)$为期望信号。

系数调节的过程为：首先自动调节滤波器系数的自适应训练步骤，然后利用滤波系数加权迟线抽头上的信号来产生输出信号，将输出信号与期望信号进行对比，所得误差值 e(k)通过一定的自适应控制算法再来调整权值，以保证滤波器处在最佳状态，达到滤波的目的。

在自适应滤波器工作时，有两种参数调整过程，这两种过程分别被命名为“学习”过程和“跟踪”过程。当不知道输入信号的统计特性时，自适应滤波器可以调整自己

参数的这一过程被称为“学习”的过程；当在处理信号过程中，输入过程的统计特性发生变化，此时自适应滤波器对自己的参数进行调整的过程被叫做“跟踪”的过程。

5.2.2 自适应滤波算法

自适应算法即是先用滤波器的输入信号、输出信号和期望信号构造一个目标函数，然后用误差序列$e(k)$按照某种准则和自适应算法，对其权系数进行调节，最终使自适应滤波器的目标函数最小化，达到最佳滤波效果。

自适应滤波器算法主要有：LMS 算法，RLS 算法等。本课题选用的是改进的 LMS 算法来进行滤波，所以 LMS 算法是本章介绍的重点。

LMS算法是基于维纳滤波理论得到的最典型的自适应算法，最早由Widrow 和 Hoff 提出[37]，是最简单也是应用最广泛的一种自适应算法。

LMS 算法最突出的优点是计算简单，运算量小，实现方便，它是一种很有用且很简单的估计梯度的方法。LMS 算法是根据最小均方误差准则进行设计的，它是用平方误差计算得到梯度矢量的，其梯度矢量为：

$$\nabla(k)=\frac{\partial e^2(k)}{\partial W(k)}=2e(k)\frac{\partial e(k)}{\partial W(K)} \tag{5-7}$$

将$y(k)=X(k)W(k)$和$e(k)=d(k)-y(k)$带入上式得到：

$$\nabla(k)=2e(k)\frac{\partial[d(k)-y(k)]}{\partial W(k)}=2e(k)\frac{\partial[d(k)-X(k)^TW(k)]}{\partial W(k)}=-2e(k)X(k) \tag{5-8}$$

权系数$W(k+1)$可简单表示为：

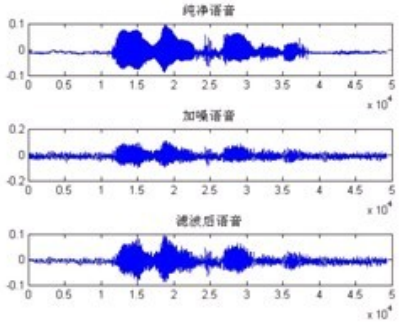
$$W(k+1)=W(k)+\frac{1}{2}\times\mu[-\nabla(k)] \tag{5-9}$$

其中μ为一个正实数，称为自适应收敛系数或步长因子。

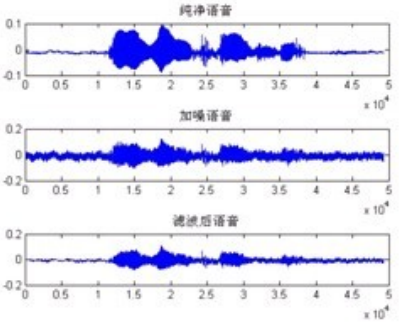
将公式（5-8）带入可得 LMS 算法最终表达式为：

$$W(k+1)=W(k)+\mu e(k)X(k) \tag{5-10}$$

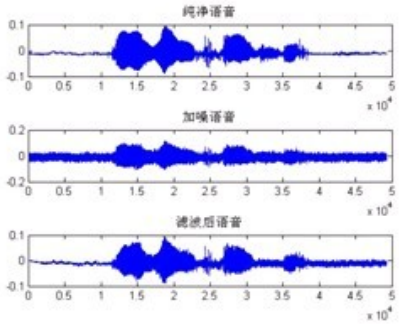
由公式（5-10）可以看出，在 LMS 算法中，下一个时刻的权系数 W(k+1)等于当前的权系数 W(k)再加上误差信号 e(k)的加权值，该加权系数为 uX(k)，可以看出，该加权值与当前的输入信号成正比，且对于权系数的所有分量来说，e(k)是相同的。



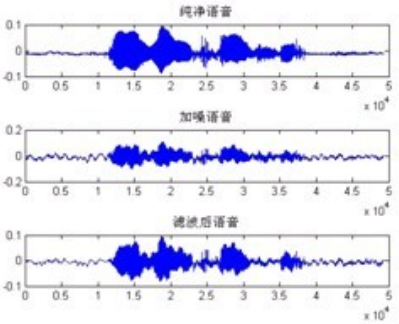
a 加入 5dB 的 tank 噪声



b 加入 5dB 的 pink 噪声

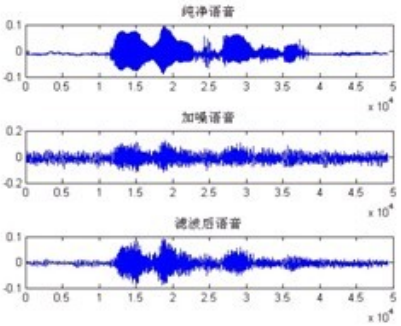


c 加入 5dB 的 white 噪声

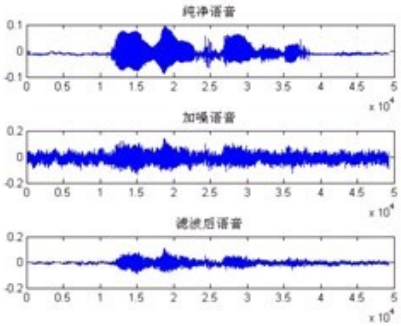


d 加入 5dB 的 volvo 噪声

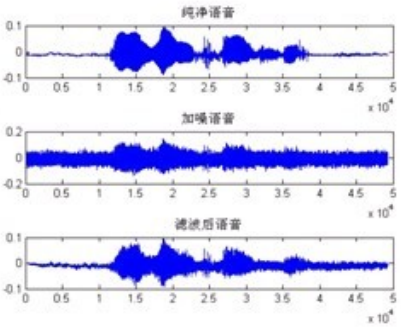
图5-3 加入SNR=5dB的噪声，滤波后的语音信号



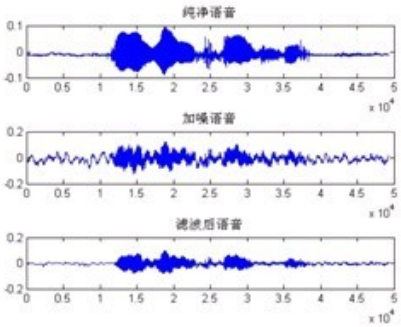
a 加入 0dB 的 tank 噪声



b 加入 0dB 的 pink 噪声



c 加入 0dB 的 white 噪声



d 加入 0dB 的 volvo 噪声

图5-4 加入SNR=0dB的噪声，滤波后的语音信号

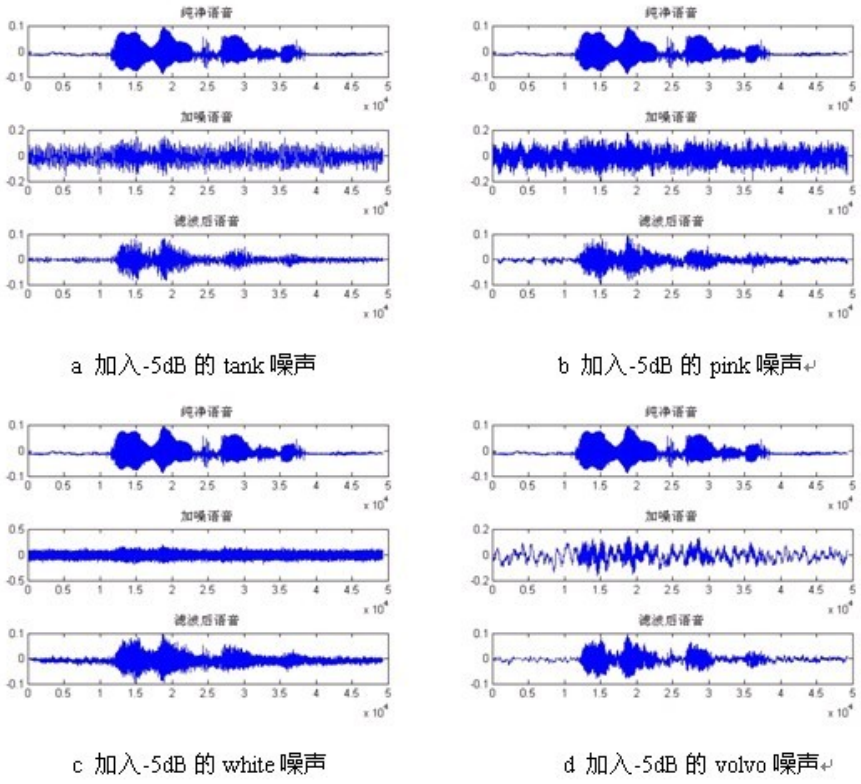


图5-5 加入SNR=-5dB的噪声，滤波后的语音信号

本文中对语音信号加入NOISEX-92噪声库中的tank、pink、white、volvo四种噪声，噪声信噪比分别选用5dB、0dB、-5dB，然后应用LMS自适应算法来对其进行降噪处理，其仿真图如图5-3、5-4、5-5所示，由仿真图可以看出，该滤波算法能达到很好的降噪效果。

第 6 章 算法实现与仿真

为了达到准确检测语音端点的目的，本文先用第五章中介绍的 LMS 自适应滤波算法对要检测的语音信号进行滤波，滤波过后再用子带自适应选择功率谱熵的语音端点检测算法进行语音端点检测，本章将详细介绍该过程。

6.1 算法过程框图

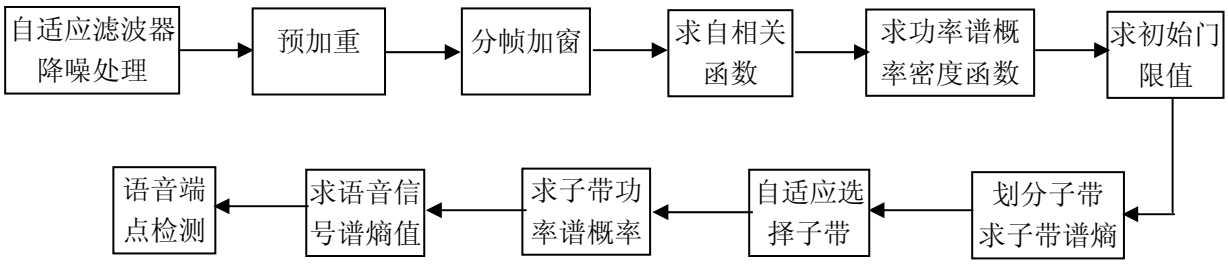


图 6-1 算法过程框图

具体步骤为：

（1）滤波去噪声。先用自适应滤波器对要检测的语音信号进行滤波。

（2）将滤波后的语音信号进行预加重处理。

（3）进行分帧加窗。在此设计中，选用窗长为 256，重叠为 128 的汉明窗，也就是指加窗之后得到的每一帧的帧长为 256，帧移为 128。

（4）对每一帧语音信号求自相关运算。

（5）对求得的每一帧的自相关函数求其 FFT 变换，从而得到每一帧语音信号中的每一点的功率谱。此处采用的是 256 点的 FFT 变换。

（6）求每一帧上的每一点的功率谱的概率。

（7）计算前 10 帧信号的平均谱熵值，以此作为初始门限值。

（8）划分子带，求子带的谱熵。

（9）计算得出最小的子带谱熵值，应用自适应选择子带方法，求出有用的子带的数目。

（10）求有用子带的功率谱概率密度函数。

（11）计算得到语音信号的谱熵值。

（12）进行语音端点检测，如果有连续的 n 帧信号的谱熵值超过门限，则判定该

段连续信号的开始处为语音起始点，如果有连续 n 帧的语音信号的谱熵值低于该门限，那么就将开始处设定为语音信号的结束位置。

（13）继续输入语音信号，应用第（12）步骤中叙述的检测方式实现语音端点检测。

6.2 仿真结果及分析

以上语音信号的仿真图都是基于 MATLAB 软件以及语音处理工具箱 VOICE BOX 仿真得到的。实验室录制的纯净语音为“语音检测”，采样率设置为 8000Hz，并选用窗长为 256 的汉明窗进行加窗、分帧处理，设置帧长为 256，帧移为 128，对于图 6-1 中的语音信号来说，总共具有 384 帧。图中对纯净语音加入的是 SNR=-5dB 白噪声信号，其加噪信号、滤波之后的信号以及检测所得的仿真结果如图 6-1 所表示：

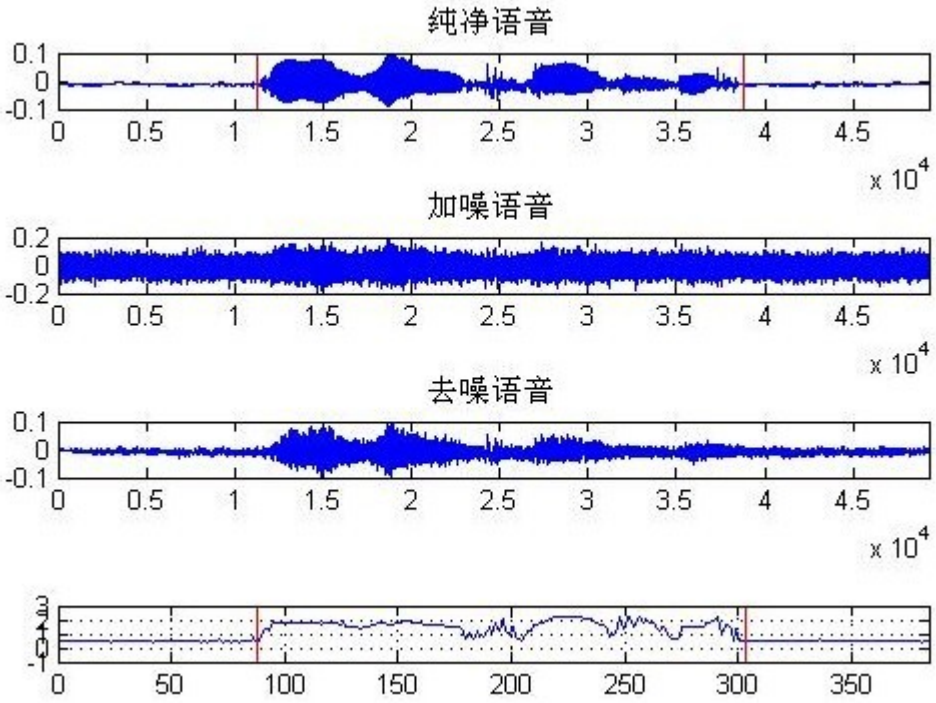


图 6-1 “语音测试”的仿真图

本文在实验室环境下录制 100 句语音，并加入 NOISEX—92 噪声库中的 4 种噪声信号（Whtie、Tank、Pink、Volvo），加入噪声的信噪比分别为 0dB、5dB、10dB、20dB，在此种条件下，分别应用短时能量及短时过零率检测法，文献[31]中的子带自适应选择谱熵检测法，以及本文中改进的检测方法进行仿真实验。在实验前，先人工标记出每个语音信号的语音端点位置，以此作为标准来判断各种检测算法的准确性。其仿真

对比结果如表 6-1 所示。

表 6-1 语音端点检测结果

噪声	加入噪声信噪比（dB）	能量及过零率检测法	子带自适应选择谱熵检测法	本文改进检测法
White	0	失效	89.8%	93.5%
	5	失效	91.6%	94.7%
	10	21.0%	94.3%	96.4%
	20	34.1%	95.6%	97.8%
Tank	0	失效	88.7%	93.6%
	5	失效	92.6%	95.2%
	10	20.4%	93.5%	96.7%
	20	32.7%	95.7%	98.1%
Pink	0	失效	87.4%	92.1%
	5	失效	91.2%	93.4%
	10	11.4%	92.8%	94.6%
	20	18.7%	93.6%	95.8%
Volvo	0	失效	90.3%	92.8%
	5	失效	91.6%	93.7%
	10	22.3%	93.5%	95.4%
	20	33.6%	94.6%	97.8%

仿真结果表明，本文应用自适应滤波技术先对语音信号降噪处理，然后再应用改进的子带自适应选择功率谱熵法来进行语音端点检测，能达到很好的检测效果。

第 7 章 DSP 硬件设计

7.1 DSP 概述

DSP 芯片即 Digital Signal Processor，它是专门进行数字信号处理的芯片，所以在诞生之初就被人们命名为 DSP—数字信号处理芯片或者称为数字信号处理器。

7.1.1 DSP 的主要结构

DSP芯片和其他处理器相比较而言，它的基本结构包括以下几个方面[38]：

（1）哈佛结构：在 DSP 芯片中，数据存储器和程序存储器是相互独立的两个存储器，各有属于自己的地址和数据总线。换句话说，也就是说数据存储器和程序存储器是存在于不同的存储空间中，我们能对其进行独立的访问。

（2）总线结构：芯片中分别存在着数据总线和程序总线。

（3）流水线结构：DSP 芯片中因为采用流水线操作，可以同时并行处理多条指令，从而使芯片的处理能力得以极大增强。

（4）专用的乘法器：乘法操作是所有程序当中必不可少的一个运算，且它决定着大多数的程序的执行快慢和复杂程度，如果处理器处理乘法的速度很快，那么处理器的运算速度就会很高。在 DSP 芯片中，采用专用的乘法器来完成乘法操作，这是 DSP 芯片优于其他处理器之处。

7.1.2 DSP 的主要特点

DSP作为数字信号处理的专用器件，其主要特点为[38]：

（1）程序空间和数据空间独立存在，在访问指令的同时也可以访问数据；

（2）发生中断时，处理速度非常快；

（3）存在硬件 I/O 口；

（4）在一个指令周期内可完成一次乘法以及一次加法运算；

（5）具有单周期内操作的多个硬件地址产生器；

（6）能够并行执行多个操作；

（7）支持流水线操作，使取指、译码、执行等操作可以并行执行；

（8）具有低开销或无开销循环以及跳转的硬件支持。

7.2 硬件结构框图

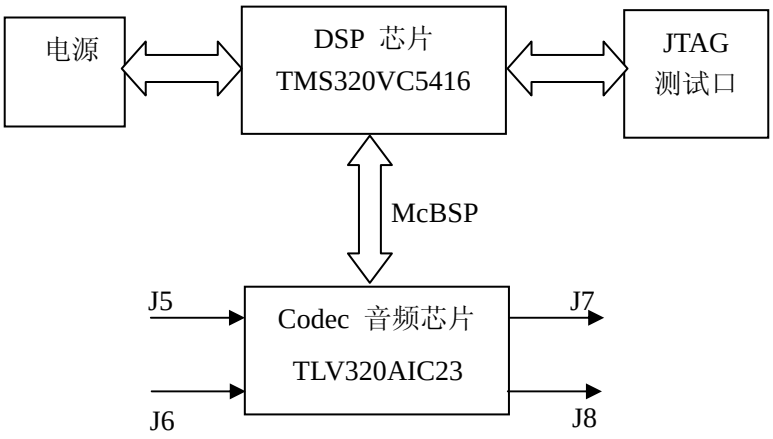


图 7-1 硬件结构框图

其中 J5 和 J6 为音频输入接口，J5 设计为单声道输入，可以直接和麦克风连接；J6 设计为立体声输入，可通过 USB 口输入音频信号。

J7 和 J8 为音频输出接口，输出的都是立体声，J8 主要用于线路输出，无增益，输出必须连接有源音箱；J7 为耳机输出，带增益，可以连接耳机，无源和有源音响等设备。

7.3 主要芯片介绍

本课题采用谱熵法来实现语音的端点检测，对其硬件系统的研究，初步设想采用 TMS320VC5416 作为主处理器并选用 TLV320AIC23 作为 Codec 芯片来实现语音端点检测的目的。

7.3.1 DSP 芯片

DSP 芯片的选择对语音端点检测系统能否完成实时的检测任务起着非常重要的作用，在选择 DSP 芯片时，要从 DSP 芯片的运算速度和存储器空间的大小上深入研究。因为语音端点检测要求实时实现，所以必须考虑运算器的运算速度是否能符合语音采样及处理的速度，另外还需注意到在进行处理时，存储器的内部资源是否够用，如果内部存储器空间不够使用时，还需要对外部存储器进行扩展。本课题中选用的是 TI 公司生产的 TMS320VC5416 作为语音端点检测系统的硬件处理器。C54x 系列 DSP 处理器是当今通信市场的主流产品，TMS320VC5416 作为 C54x 系列的第三代芯片，以低功耗和高性能而著称，它在运算速度和存储空间等各个方面都能达到该课题的要求，

以下是对该芯片的介绍。

TMS320VC5416(以下简称VC5416)是TI公司的一款16bit定点高性能DSP, 它的中央处理单元具有改进的哈佛结构、低功耗设计和高度并行性等特点。它的主要性能有[39]：

- 速率很高，能达到160MI/s；
- 包括3条16位的数据存储器总线以及1条程序存储器总线；
- 包括1个40位的桶形移位寄存器和2个40位的累加器；
- 1个17×17的乘法器和1个40位的专用加法器；
- 最大8M×16bit的扩展寻址空间；
- 内置128k×16bit的RAM和16k×16bit的ROM；
- 3个多通道缓冲串口(McBSP)；
- 配有PCM3002, 可对语音进行A/D和D/A 转换。

由于VC5416功耗低，性能高，其分开的数据和指令空间使该芯片具有高度的并行操作能力，在单周期内允许指令和数据同时存取，再加上高度优化的指令集，使得该芯片具有很高的运算速度并且该芯片本身具有丰富的片内存储器资源和多种片上外设，因此在工程界得到广泛应用，尤其是在语音编码和通信应用方面。本课题中语音检测算法采用该DSP芯片实现，在运算速度上能满足算法要求，进而满足实时性的标准。

7.3.2 音频芯片

TLV320AIC23是TI公司的一款高性能、集成有模拟功能的立体声音频Codec芯片。它的主要特性有[40]：

- 内置耳机输出放大器，支持MIC 和LINE IN 两种输入方式(二选一)，且对输入和输出都具有可编程增益调节；
- 芯片中的A/D转换器和D/A转换器采用过取样数字内插滤波器的多位Sigma-Delta技术，并且A/D转换器和D/A转换器都高度集成在芯片内部；
- 因为采用了Sigma-Delta过采样技术，所以数据传输字长可以为16位、20位、24位和32位，采样率可以设为8kHz～96kHz；
- AIC23还具有低功耗的特点，在回放模式下功率只有为23mW，省电模式下小于15uW；占用面积也很小，只有25mm^{2}的面积；
- A/D转换器和D/A转换器的信噪比可以达到90db和100db；

- 它的核心数字电压为1.42V-3.6V的，兼容C54x系列的DSP内核电压；
- 具有2.7V-3.6V的缓冲器和模拟，与C54x系列的DSP内核电压兼容；
- 软件控制通过TMS320VC5416的McBSP接口；
- 音频数据输入输出通过TMS320VC5416的McBSP接口。

由此可知AIC23有体积小，成本低廉以及高性能等优点，因此它成为可移动的数字音频播放和录音产品中的模拟输入输出应用系统的理想选择。

7.4 McBSP 接口

VC5416 的DSP芯片提供三个高速、全双工的、多通路缓冲串行口。它的英文简写为McBSP（Multichannel Buffered Serial Port），分别为McBSP0、McBSP1、McBSP2。McBSP可直接与C54x系列的DSP芯片、音频芯片或者系统中的其他设备相接口。它的设计是基于其他 54x系列的标准串口扩展得到的，其特点如下[39]：

- 全速双工通信；
- 双缓存数据寄存器，可使连续传送的数据流；
- 接收和发送数据都是适用独立的帧和时钟；
- 多通道发送和接收数据，通道数能达到 128 个；
- $\mu-L$aw 和 A-Law 的数据压缩和扩展形式；
- 接收和发送的字的长度的范围包括：8bit、12bit、16bit、20bit、24bit 和 32bit；
- 帧同步信号和数据时钟信号的极性是可编程控制的；
- 内部可编程的时钟信号和帧信号发生器。

另外，McBSP 能直接与下列格式接口：

- T1/E1 帧；
- 与 MVIP 相兼容；
- ST-BUS 从器件；
- AC97 从器件；
- IOM-2 从器件；
- IIS 从器件；
- 串行设备接口—SPI 器件。

McBSP 包括一个数据通道和一个控制通道，它的组成分为引脚的接收发送部分、

时钟信号和帧同步信号产生器、多通道选择以及 CPU 中断信号和 DMA 同步信号。通过六个引脚，分别是串行数据接收引脚 DR，串行数据发送引脚 DX，接收帧同步引脚 FSR，发送帧同步引脚 FSX，接收时钟引脚 CLKR、发送时钟引脚 CLKX 来将数据和控制通道连接到外部设备。

McBSP与外界进行数据交换，主要是通过DX引脚发送数据，通过DR引脚来接收数据，DSP的CPU或者DMA从数据接收寄存器DRR[1/2]读取接收到的数据，发送时，把数据发送到数据发送寄存器DXR[1/2]。当数据写入到DXR[1/2]之后，通过传输移位寄存器XSR[1/2]，输出到DX引脚上；从DR引脚上接收到数据后，通过接收移位寄存器RSR[1/2]，移位存储到RSR[1/2]上，并复制到接收缓存寄存器RBR[1/2]上。然后，通过RBR[1/2]复制到DRR[1/2]，此时可由CPU或者DMA读出[38]。这就是在McBSP内部进行数据接收与发送的全部过程。

McBSP 控制寄存器可以被 CPU 访问，控制通道的任务主要有内部时钟、帧同步信号的产生和控制和多通道的选择控制，另外控制通道还负责传送一些状态和事件信息给 CPU 或者 DMA，这些信息包括 2 个中断信号和 4 个事件信号。

McBSP 的接收时钟 CLKR 和发送时钟 CLKX 可以由外部设备提供也可以由内部时钟产生器产生；帧同步信号 FSX 和 FSR 的输入和输出极性也可编程选择；串口的信号发送和接收部分既可单独运行又可以合在一起配合工作。

在 SPI 模式下，系统通常有一个主设备和一个或多个从设备组成，接口包括以下四个信号：串行数据输入(也称为主进从出，或 MISO)；串行数据输出(也称为主出从进，或 MOSI)；串行移位时钟(也称为 SCK)；从使能信号(也称为 SS)。McBSP 的时钟停止模式与 SPI 协议兼容，当 McBSP 处于时钟停止模式时，发送器和接收器是内部同步的，可以将 McBSP 作为 SPI 主设备或从设备。其中，可将发送数据帧时钟(FSX)用作从使能(即 SS)，而将发送数据位时钟(CLKX)用作 SPI 协议中的 SCK。由于接收数据位时钟和接收数据帧时钟在内部与 FSX 和 CLKX 是相连的，因此，该管脚不能用于 SPI 模式。本文中将 TMS320VC5416 的 McBSP 配置为时钟停止模式，串口接收控制寄存器 SPCR1 的时钟停止模式位 CLKSTP 和串口引脚控制寄存器 PCR 的发送时钟极性位 CLKXP 配置为 CLKSTP＝11，CLKXP＝1(时钟开始于下降沿，有延时)。

7.5 TLV320AIC23 与 TMS320VC5416 的连接图

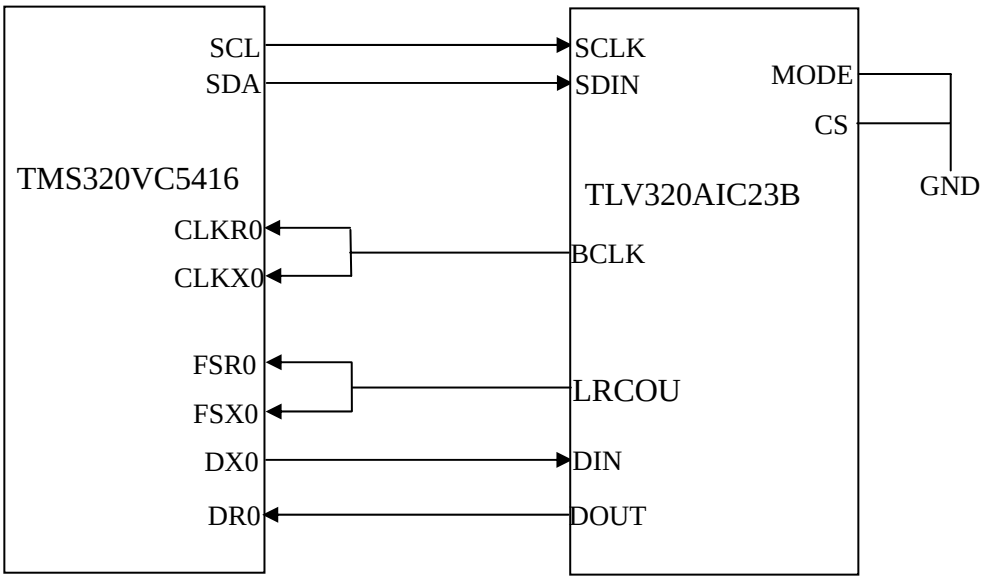


图 7-1　VC5416 与 AIC23 的连接图

SCLK为控制端口移位时钟，适用于SPI和I^2C两种模式。

SDIN 为控制端口串行数据输入。用来传输配置 AIC23 内部寄存器的数据。

BCLK为I^2S串行数据传输时钟。当AIC23 为主模式时，BCLK由AIC23 产生并提供给DSP，频率为主时钟的 1/4；从模式时，由DSP产生。

LRCOUT为I^2S格式数据输出帧同步信号。

DIN为I^2S格式串行数据输入端，送入立体声DAC。

DOUT为I^2S格式串行数据输出端，由立体声ADC产生。

MODE为串行接口输入模式选择端。0 为I^2C模式。1 为SPI模式。

CS为控制端口输入和地址锁存选择端。在SPI控制模式下，作为数据锁存控制端。在I^2C控制模式下，定义外设的 7 位地址。

把DSP的CLKR0 和CLKX0 两个信号接在一起，FSX0 和FSR0 两个信号也接在一起，这样就达到了使同步串行接口的接收和传送两边都是同样的时钟信号和帧同步信号。DSP通过I^2C总线对AIC23 进行控制设置，用McBSP0 口来进行串行数据收发。

TLV320AIC23B的工作模式共有 4 种，分别为右声道排列模式、左声道排列模式、I^2S模式、DSP模式，由AIC23 的数字音频格式控制器中的FOR0 和FOR1 位来控制。在这里采用 DSP 模式，将 FOR0 和 FOR1 设置为 11。

AIC23B 设为主模式，它的主时钟可由 12MHz 的 USB 时钟来提供，采样输入长度为 16bit，采样率为 8KHz。McBSP 的接收时钟与帧同步信号都由 AIC23B 来提供。

第 8 章 结束语

8.1 总结

本课题所研究的语音信号的端点检测是语音信号处理中非常重要的一项预处理技术，有着广泛的应用价值。

本文所做的工作主要有以下几个方面：

（1）介绍了语音信号处理中的一些基础处理知识，例如短时分析技术、预加重、加窗和分帧等。

（2）对多种传统的语音端点检测方法——短时能量检测法、过零率检测法等进行了介绍和对比，陈述了各种方法的优缺点。

（3）通过分析语音信号以及噪声信号的特点，选取稳健性很好的谱熵来作为进行语音端点检测的特征参数。

（4）对子带自适应选择技术进行了详细分析，并对此进行了改进。

（5）把基于子带自适应选择的功率谱熵端点检测算法与自适应滤波技术相结合，达到了先对信号进行滤波，提高语音信号的信噪比之后再进行语音端点检测的目的，仿真证明该方法能达到很好的检测效果。

（6）研究语音信号检测的硬件系统，给出了其 DSP 芯片与语音芯片的连接示意图。

8.2 下一步的工作

虽然本文介绍的语音端点检测算法在仿真中已经取得了很好的成效，但是随着人们对其检测准确性的要求的提高，特别是为达到简便、快速、实时和稳健性好的要求，该算法还应进行进一步的研究。

另外，本文的 DSP 硬件实时检测系统还未设计完成，还需要进一步的研究实现。

参考文献

[1] 胡航.语音信号处理第 3 版[M].哈尔滨:哈尔滨工业大学出版社,2005.

[2] 韩韬.基于强背景噪声下的语音端点检测算法及实现[D].湖南:湖南大学,2007.

[3] 沈宏余,李英.语音端点检测方法的研究[J].科学技术与工程,2008,(8) :4396－4405.

[4] 武光利,戴玉刚,马宁.基于短时平均幅度和短时平均过零率的藏语语音端点检测研究[J].福建电脑,2007,(3):116–122.

[5] 陈四根.基于熵函数的语音端点检测方法[J].声学与电子工程,2001,(1):28-30.

[6] 唐永锋,霍春宝.噪声环境下语音信号端点检测算法的研究与改进[J].人工智能及识别技术,2007,1386–1387.

[7] 段红梅,汪军等,马良河等. 隐马尔可夫模型在语音识别中的应用[J].工科数学,2002,18(3):16–20.

[8] 蔡魁杰. 基于支持向量机的汉语语音端点检测和声韵分离[D].哈尔滨:哈尔滨工程大学,2007.

[9] 乔峰.基于信息熵和神经网络的语音端点检测算法研究[D].太原:太原理工大学,2007.

[10] 赵志诚.DSP 技术的发展及应用[J].仪表技术与传感器,1999:1–2.

[11] 张雄伟,陈亮,杨吉斌.现代语音处理技术及应用[M].北京:机械工业出版社,2003.

[12] 胡广书.数字信号处理——理论、算法与实现[M].北京:清华大学出版社,2003.

[13] 焦卫东,杨世锡,钱苏翔等.乘性噪声消除的同态变换盲源分离算法[J].浙江大学学报(工学版),2006,40(4):581–584

[14] He Suning,Yu Juebang.A Novel Chinese Continuous Speech Endpoint Detection Method Based on Time Domain Features of the Word Structure [J].IEEE Int.Conf.on Commun.Circuits and Systems and West Sino Expositions,2002:992–996.

[15] 樊昌信.通信原理[M].北京:国防工业出版社,2001.

[16] 李昱,林志谋,黄云鹰等.基于短时能量和短时过零率的 VAD 算法及其 FPGA 实现[J].电子技术应用,2009,9:110–113.

[17] 杨晓玲.基于 DSP 的实时语音检测系统的研究[D].武汉:华中师范大学,2006.

[18] 国雁萌,盛任农,牟英良.基于能量和浊音特性的语言端点检测[J].计算机工程与应用,2006,43.

[19] 胡光锐.基于倒谱特征的带噪语言端点检测算法[J].电子学报,2000,28(1):95–97.

[20] 唐永锋,霍春宝.噪声环境下语言信号端点检测算法的研究与改进[J].人工智能及识别技术,2007:1386–1387.

[21] 果永振,何遵文,刘畅等.基于 DSP 实现语音端点检测[J].华北科技学院学报,2003,3:46–48.

[22] 李洪波,于洪志.噪声环境下语音识别的端点检测技术[J].西北民族大学学报(自然科学版),2007,28(65).

[23] 刘华平,李昕,郑宇等.一种改进的自适应子带谱熵语音端点检测方法[J].系统仿真学报,2008,20(5):1366–1371.

[24] 傅祖芸.信息论—基础理论与应用(第二版) [M].西安:电子工业出版社,2007.

[25] Shen J L, Hung J W, Lee L S.Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments [C]// Processing.Sydney ICSLP (S0160-5840),Sydeny,Australia,Nov-Dec 1998,CD2ROM,1998:32–235.

[26] 严剑峰,付宇卓.一种新的基于信息熵的带噪语音端点检测方法[J].计算机仿真,2005,22(11):117–139.

[27] Hemant Misra,Shajith Ikbal.Multi-resolution spectral entropy feature for robust ASR [C].ICASSP－05 (S1053-5888).Philadelphia,PA,2005,(3):253–256.

[28] 侯周国,钱盛友,姚畅.短时域语言端点检测中谱熵算法的改进[J].计算机工程与应用,2006:55–56.

[29] 姚屏.基于 DSP 的语音检测与神经网络自适应滤波研究[D].长沙:中南大学,2005.

[30] Gin-Der Wu,Chin-Teng Lin.Word Boundary Detection with Mel-Scale Frequency Bank in Noisy. Environment [J].Ieee Transactions on Speech and Audio Processing (S1063-6676), 2000, 8(5): 541–554.

[31] Bing-Fei Wu,Kun-Ching Wang. Roust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments[J].IEEE Transactions on Speech and Audio Processing (S1063-6676),2005,13(5):762–775.

[32] C.T.Lin,J.Y.Lin,G.D.Wu.A Robust Word Boundry Detection Algorithm for Variable Noise-level Environment in cars[J].IEEE Transactions Intelligent Transportation System,2002,3:89–101.

[33] 李金宝,屈百达,徐宝国等.基于自适应子带功率谱熵的语音端点检测算法[J].计算机工程与应用,2007,43(12):57–65.

[34] Simon Haykin,Adaptive Filter Theory [M].Beijing:Publishing House of Electronics Industry,2004.

[35] 苏剑峰,吴海涛,边玉敬.自适应滤波技术的研究[J].飞行器测控学报,2006,25 (2) :71–74.

[36] 樊殊昱.LMS 算法的改进研究及其在语音增强方面的应用和性能评估[D].西安:电子科技大学,2007.

[37] B.Widrow,S.D.Stearns.Adaptive Signal Processing.Prentice Hall,Englewood Cliffs,NJ,1985.

[38] 彭启琮,李玉柏,管庆.DSP 技术的发展与应用[M].北京:高等教育出版社,2002.

[39] TMS320VC5416 Fix-Point Digital Signal Processor Data Manual[Z].March 1999-Revised October 2008.

[40] TLV320AIC23 Data Manual.Texas Instuments[Z].July,2001.

致 谢

首先我要衷心感谢我的导师龙海南教授。感谢他对我三年来的悉心照顾，使我不仅在学习中受益匪浅，而且在生活中使我懂得如何与人更好的相处。另外龙老师认真严谨的学术作风以及对待问题一丝不苟的探索精神给我留下了尤为深刻的印象。此外，感谢龙老师给予我很大的发展空间，感谢龙老师引导我在生活以及相关的研究领域获取更多的知识。感谢他三年来对我的关心和爱护！

感谢通信电路研究实验室的兄弟姐妹，感谢师姐、师哥、师弟、师妹对我学习上的倾囊辅导，感谢他们对我的生活上的关心和爱护。在实验室度过的每一天都是我这一生中最美好的回忆，感谢他们使我的这三年研究生生涯过的无比充实与快乐。

感谢 07 级通信与信息系统专业的所有同学，感谢他们这三年来对我生活上和学习上的无私帮助，感谢他们为我营造了一个健康温馨、积极乐观的班级环境。

我还要感谢我的家人和我的朋友，感谢多年来他们对我的支持、鼓励和信任，感谢他们对我的一切的无私奉献。

最后，我要感谢在白忙之中抽出时间来评阅本文的所有评委，感谢你们为我细心的评阅文章，谢谢！

攻读学位期间取得的科研成果

[1] 龙海南，张翠改．数字温度计和温控器 DS7505 及其应用[J]．电子设计工程，2009，17（1）：106－107．

[2] 龙海南，张翠改．An Improved Method for Robust Speech Endpoint Detection．ICMLC 2009．