

语音信号的基音频率提取算法研究

1 前言

基音是指发浊音时声带振动所引起的周期性,而声带振动频率的倒数就是基音周期。基音周期具有时变性和准周期性,它的大小与个人声带的长短、厚薄、韧性和发音习惯有关,还与发音者的性别、年龄、发音时的力度及情感有关,是语音信号处理中的重要参数之一,它描述了语音激励源的一个重要特征。基音周期的估计称为基音检测(Pitch Detection),基音检测的最终目标是找出和声带振动频率完全一致的基音周期变化轨迹曲线,如不可能则找出尽量相吻合的轨迹曲线。

然而由于人的声道的易变性及其声道特征的因人而异,而基音周期的范围又很宽,且同一个人在不同情态下发音的基音周期也不同,加之基音周期还受到单词发音音调的影响,故实际中的基音周期的精确检测是一件比较困难的事情。然而,尽管语音信号的基音检测有许多困难但由于它在语音信号处理中的重要作用,促使广大学者争相涉足该领域,提出了各种各样的基音检测算法。

2 选题背景

2.1 基音检测技术的研究进展

早在 70 年代, L. R. R 等人就进行了自相关函数法检测语音信号的基音周期的研究工作,它是一种时域上的基音检测算法,算法的精确性高,计算量不大,是目前各种应用中最为常用的基音检测算法。

1967 年, A. M. Noll 提出用倒谱法(Cepstrum)检测语音信号的基音周期。这是一个频域上的检测算法,这种方法检测基音周期精确度很高,抗噪性能好,主要的缺陷是计算量太大,要用到傅立叶变换和对数运算,不利于实现。

1972 年, J. D 等提出简单逆滤波追踪法(SIFT)检测语音信号的基音周期,这是一种时域和频域相结合的算法,是一种精确度和计算量较为折中的算法,它利用逆滤波去除声道共振峰的影响,使基音信息更为突出。

1974 年, M. J. Ross 等人提出平均幅度差函数法检测语音信号的基音周期,这是一种时域上的算法,也是最简单的基音检测算法,它只需在时域上进行简单的加减和少量的除法运算,运算量很小,但是很容易产生半基音和倍基音,目前还有很多人在

不断的提出改进的 AMDF 算法。

到 1976 年, L. R. R 等人系统总结了之前的各种语音信号的基音检测算法, 并进行了全面的比较。这些算法都是假定语音信号在一帧内是平稳的且一帧内包含两个以上的基音周期, 所以它们不能很好反映语音信号的时变特性, 而且只能求出一帧内的平均周期。

1992 年, S. K 等人最早提出了基于小波变换的基音检测算法, 它是一种利用变换的基音检测算法, 通过小波变换后的幅度和相邻两个尺度下的小波变换的局部最大值是否一致来进行清浊音判断, 浊音的小波变换的局部最大值点即为 G_0 , 而相邻两个 G_0 的距离即为语音信号的基音周期。受小波变换的基音检测算法的启发, 不断地有人提出了各种改进的小波变换基音检测算法, 如: 小波变换偏移补偿的基音检测算法, 利用小波变换和其它方法结合的基音检测算等。

2.2 基音检测技术的研究现状

从国内外研究现状来看, 基音检测技术的研究热点和难点已经集中于处理低信噪比语音。着眼于基音的检测方法, 主要有以下三个方面的研究: (1)稳定并提取准周期性信号的周期性方法; (2)因周期混乱, 采取基音提取误差补偿的方法; (3)消除声道噪声影响的方法。人们从语音信号的时域特性、频域和时一频混合特性三个方面出发, 已经开发了许多基音检测方法, 这些方法中的一些方案已经得到了应用。基音检测方法大致上可以分为三类: (1)时域估计法, 直接由波形来估计基音周期, 常见的有: 自相关(ACF)法^[31]、平均幅度差法^[32]等; (2)频域估计法, 利用同态分析方法将声道的影响消除, 得到属于激励部分的信息, 然后求取基音周期, 常见的有: 谐波积谱(HPS)法, 简化逆滤波(SIFT)法^[33], 倒谱(cepstrum)法^[34]等; (3)混合法, 基于以上方法的衍生组合算法, 如: 自相关法和平均幅度差法相结合。这些方法在实验室特定条件下都取得了不错的效果。

表 1 列出了几种典型的基音检测方法及其特征^[30]

表 1 几种典型的基音检测方法及其特征

分类	基音提取法	特征
波形估计法	并行处理法	由多种简单的波形峰值检测器决定提取的多数基音周期。
	数据减少法	根据各种理论操作，从波形去掉修正基音脉冲以外的数据。
	过零数法	关于波形的过零数，着眼于重复图形
相关处理法	自相关法	语音波形的自相关函数，根据中心削波，平坦处理频谱，采用峰值削波可以简化运算。
	SIFT 法	语音信号波形降低取样后，进行 LPC 分析，用逆滤波器平坦处理频谱，通过预测误差的自相关函数，恢复时间精度。
	AMDF 法	采用平均幅度差函数检测周期性，根据预测误差信号的 AMDF 也可以进行提取。
变换法	倒谱法	根据对数功率谱的傅里叶逆变换，分离频谱包络和细微结构。
	循环直方图法	在频谱上，求出基频高次谐波成分的直方图，根据高次谐波的公约数决定基音。

目前，基音检测主要存在以下困难^[35]：

(1)语音信号变化十分复杂，声门激励的波形并不是一个完全的周期序列。在语音的头尾部并不具有声带振动那样的周期性，对有些清浊音的过渡帧很难判定它属于周期性还是非周期性，从而对估计基音周期带来一定的影响。

(2)从语音信号中去除声道的影响，直接取出仅与声带振动有关的声源信息并非易事。如声道共振峰可能强烈改变声门波形的结构，从而严重影响激励信号的谐波结构，会给基音检测造成困难。

(3)语音信号是准周期的，且共振峰结构和噪声有时会影响波峰和过零率，很难准确定位基音周期的开始和结束。

(4)区分清音语音和低电平语音是导致基音检测困难的另一个重基音检测算法研究及其在方言辨识中的应用要因素。在许多情况下，清音语音与低电平浊音段之间的过度段是非常细微的，确认它是极其困难的。

(5)在实际应用中，背景噪声强烈影响基音检测的性能。

(6)基音频域变化范围大，从老年男性的 80Hz 到儿童女性的 500Hz，接近三个倍频程，给基音检测带来了一定的困难。由于这些困难，尽管基音提取的方法很多，但迄今尚未找到一个完善的方法可以对各类人群包括男人、女人、老人、小孩及不同语种和各种环境条件情况下都能获得满意的检测结果。

3 方案论证

3.1 语音信号处理基础

3.1.1 语音信号产生

3.1.1.1 发音器官

人类的语音是由人体发音器官在大脑控制下的生理运动产生的。人体发音器官由三部分组成：肺和气管，喉，声道^[3,21]。

肺是胸腔内的一团有弹性的海绵状物质，它可以储存空气。其主要生理功能是使血液和空气之间进行气体交换，即将空气中的氧气吸入血液，而将血液中的二氧化碳排入空气，这就是人体的呼吸功能。在说话时，为了保持语音有一定程度的连续性，人的呼吸就不得不有短暂的停顿，其特点是吸气短，呼气长，且呼吸受到句子结构的控制，并没有一个固定的规则。空气由肺部排入喉部，经过声带进入声道，最后由嘴辐射出声波，形成语音。

喉是一种由软骨和肌肉组成的复杂系统，其中有重要的发音器官-声带（vocalcords）。它主要包括环状软骨、甲状软骨、钩状软骨和声带。声带受到喉部软

骨和肌肉的控制。当声带分开时，处于正常呼吸状态。两片声带之间的空间叫做声门 (glottis)。当声带闭合时，肺道便被封闭起来构成一个密闭的小室。

声带每次开启和闭合一次的时间就是基音周期 (pitch period), 它的倒数就是基音频率 (pitch frequency)。基音频率取决于声带的大小、厚度、松紧程度以及声门上下之间的气压差的效应等。一般基音频率越高，声带被拉得越长、越紧、越薄，声门的形状也变得越细长，而且这时声带在闭合时也未必是完全闭合。基音频率最低可达到 80Hz 左右，最高可达 500Hz 左右。其范围随发音人的性别、年龄及具体情况而定。老年男性偏低，小孩和青年女性偏高。

声道是它是一根从声门延伸至口唇的非均匀截面的声管，其外形变化是时间的函数。声道是气流自声门、声带之后最重要的、对发音起决定性作用的器官，它发出不同音时其形状变化是非常复杂的。成年男子声道的平均长度约 17cm，而声道的截面积取决于其发音器官的位置。发音过程中声道的截面积由舌头、唇、上腭、小舌等的位置决定。

3.1.1.2 语音产生过程

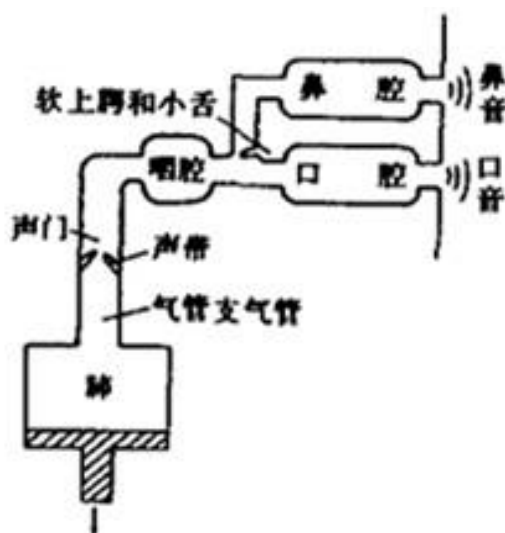


图 1 语音产生机理示意图

图 1 为语音产生机理示意图。在发音过程中，肺部与相连的肌肉相当于声道系统

的激励源。当声带处于收紧状态时。流经的气流使声带振动，这时产生的声音称为浊音（voiced sound），不伴有声带振动的音成为清音（unvoiced sound），当声带处于放松状态时，有两种方式能发出声音：一种方式是通过舌头，在声道的某一部分形成狭窄部位，也称为为收紧点，当气流经过这个收紧点时会产生湍流，形成噪声型的声音。这时对应的收紧点的位置不同及声道形状的不同，形成不同的摩擦音另一种方式是声带处于松懈状态，利用舌头和嘴唇关闭声道，暂时阻止气流。当气流压力非常高时，突然放开舌与唇，气流被突然释放产生短暂脉冲音。对应于声道闭紧点的不同位置和声道的形状，形成不同的爆破音。

为了发出各种各样的声音，需要调整声道的形状，称之为调音（articulation）。声道各部位的动作称为调音运动（articulation movement）。调音用的声道的各部分器官叫调音器官（articulation organ），包括舌、腭、唇和嘴等声道中可以自由活动的部分。在调音器官中，因调音而产生的声道固定部位的狭窄位置称为调音点（place of articulation）。声带的状态，包括它的位置、形状、各个不同的调音器官的大小等随时间的变化产生不同音色的语音。这是因为对应不同的声道形状就有不同的传递特性。

3.1.2 语音信号的数字化

3.1.2.1 语音信号的采样与量化

语音信号是时间和幅度都连续变化的一维模拟信号，要想在计算机了对它进行处理，就要先进行采样和量化，将它变成时间和幅度都离散的数字信号。如下图：

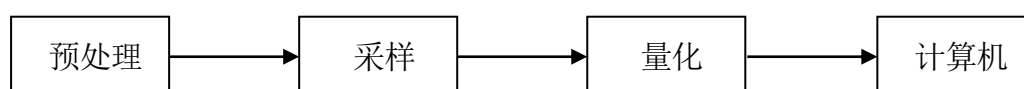


图 2 语音信号数字化框图

所谓采样就是把模拟信号在时间域上进行等间隔取样，其中两样本之间的间隔称为采样周期，它的倒数称为采样频率。根据采样定理，当采样频率大于信号最高频率的两倍时，在采样过程中就不会丢失信息，并且可以根据采样后的信号重构原始信号。实际的信号常有一些底能量的频谱部分超过采样频率的一半，如浊音的频谱超过 4KHz 的分量比其峰值要低 40dB 以上；对于清音，即使超过 8KHz，频率部分也没有明显的下降，因此语音信号所占的频率范围可达 10KHz 左右。在实际的语音信号处

理中，采样频率一般为 8KHz~10KHz。

采样后的信号在时间域上是离散的形式，但在幅度上还保持着连续的特点，所以要进行量化，量化的目的是将信号波形的幅度值离散化。一个量化器就是将整个信号的幅度值分成若干个有限的区间，并且把落入同一个区间的样本点都用同一个幅度值表示，这个幅度值称为量化值。

3.1.2.2 语音信号的短时加窗处理

经过数字化的语音信号实际上是一个时变信号，为了能用传统的方法对语音信号进行分析，应假设语音信号在 10ms-30ms 短时间内是平稳的。后面所有的分析都是在语音信号短时平稳这个假设条件进行的。

为了得到短时的语音信号，要对语音信号进行加窗操作。在加窗的时候，不同的窗口选择将影响到语音信号分析的结果。在选择窗的时候一般有两个问题要考虑。

① 窗函数的形状

窗函数可以选用矩形窗，即：

$$W(n) = \begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & \text{其他} \end{cases} \quad (3-1)$$

或其他形式的窗函数，例如汉明窗（Hamming），即：

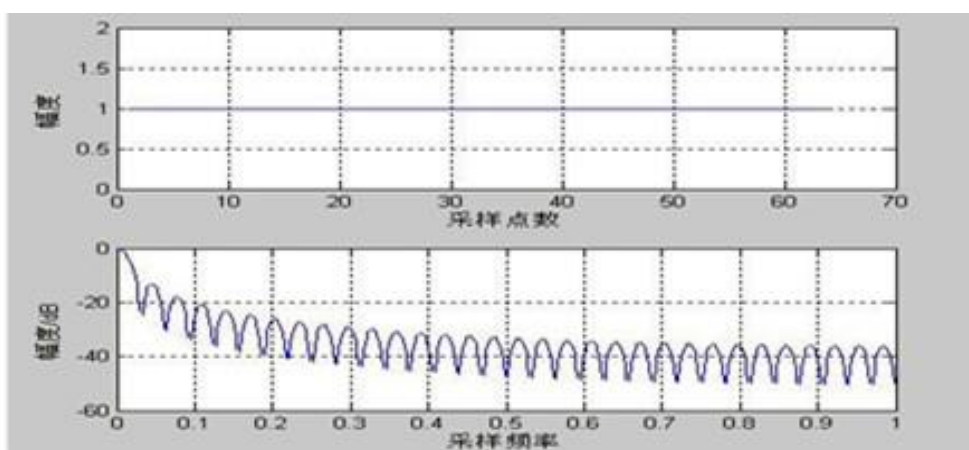
$$W(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (N-1)) & 0 \leq n \leq N-1 \\ 0 & \text{其他} \end{cases} \quad (3-2)$$

式中，N 为窗口长度。

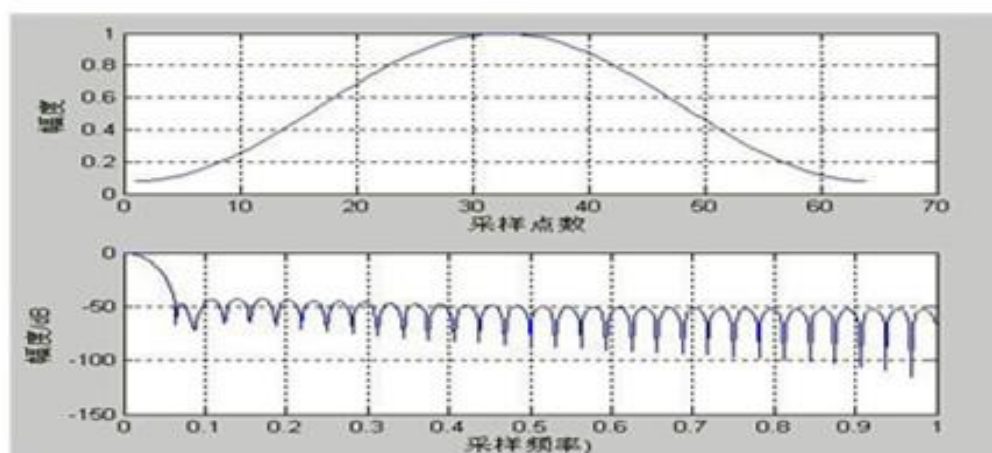
虽然这些窗函数的频率响应都具有低通特性，但不同的窗函数形状将影响分帧后短时特征的特性。下面以矩形窗和汉明窗为例对窗函数进行比较。

如图中给出了在 N=64 时的矩形窗和汉明窗及其频率响应对应的对数幅度。可以看到，汉明窗的第一个零值频率位置比矩形窗要大一倍左右，即汉明窗的主瓣大约

是同样宽度矩形窗的两倍。同时也可以很明显地看到，在通带外汉明窗的衰减较相应的矩形窗大得多。



(a) 矩形窗



(b) 汉明窗

图3 两种常用窗函数时域及频率幅度图

由图可知，矩形窗的旁瓣太高，会产生严重的泄漏现象，因此只在某些特殊场合中采用。汉明窗旁瓣较低，可以有效地克服泄漏现象，具有更平滑的低通特性，因此

应用最为广泛^[7]。

② 窗函数的长度

不论什么样的窗函数，窗函数的长度对能否反映语音信号的幅度变化起决定性的作用。如果 N 特别大，即等于几个基音周期量级，则窗函数等效于很窄的低通滤波器，此时信号短时信息将很慢地变化，因而也就不能充分地反映波形变化的细节；如果 N 太小，滤波器的通带变宽，则不能得到较为平滑的短时信息，因此窗函数的长度要选择合适的。窗函数的衰减基本上与窗函数的持续时间无关，因此当改变窗函数的长度 N 时，会使带宽发生变化^[3]。

前面的窗函数长度是相对于语音信号的基音周期而言的。通常认为一个语音帧内，应含有 1~7 个基音周期，然而不同人的基音周期变化范围很大，基音周期持续时间会从高音调的女性或儿童的约 20 个采样点（采样频率为 10kHz）变化到很低音调的男性的 250 个采样点，这意味着在进行分析时可能需要多个不同的 N 值，所以 N 的选择比较困难。通常在采样频率为 10kHz 的情况下， N 的选择在 100~200 量级是合适的（也就是说 10ms~20ms 持续时间）。

对语音信号的时域分析来说，窗函数的形状是非常重要的，矩形窗的谱平滑性较好，但波形细节丢失，并且矩形窗会产生泄漏现象，而汉明窗可以有效地克服泄漏现象，具有更平滑的低通特性，应用最为广泛^[7]。

3.2 语音信号的特征分析

3.2.1 语音信号的时域分析

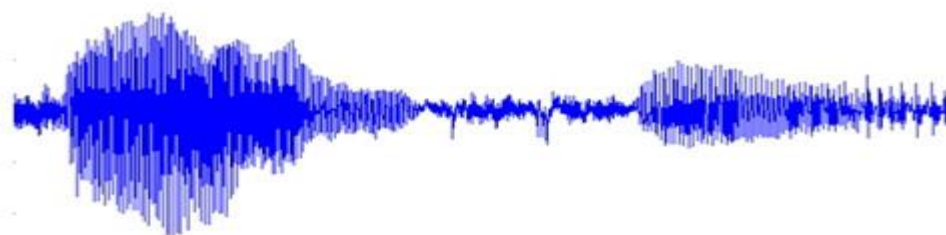
对信号分析最直接的方法是以时间为自变量进行分析，语音信号典型的时域特征包括短时能量、短时平均过零率、短时自相关系数和短时平均幅度差等。在这一节中，主要对短时能量、短时平均过零率等特征及它们的应用加以介绍。

3.2.1.1 语音信号的时域表示

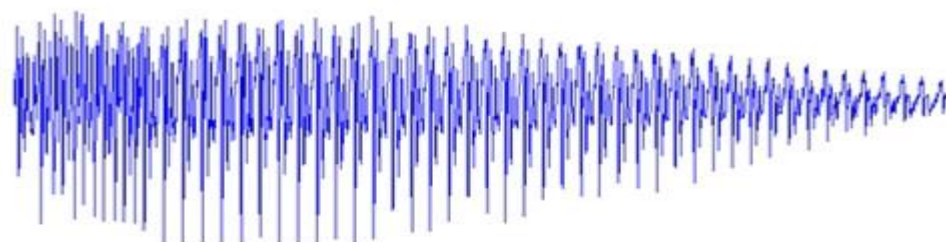
在进行语音信号数字处理时，最先接触、最直观的是它的时域波形。通常是将语音信号转换成电信号，再用 A/D 转换器将其转换成离散的数字信号，然后存入计算机内存中^[2]。

下图 a 是一个女性说“开始”的时域波形，语音数据是在实验室环境下用普通麦

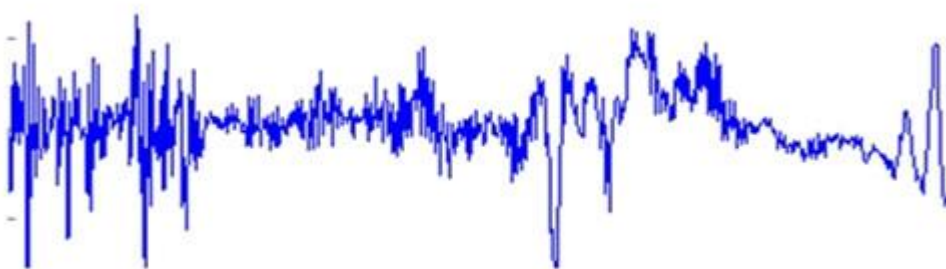
克风录制的。采样频率为 8000Hz，采样后进行 16 位量化。图中横轴为时间，纵轴表示信号的幅度。从图中可以看出，虽然无法辨别语音波形的细节，但可以看出语音能量的起伏，以及语音信号随时间变化的过程。图 b 是将“开”的元音部分/a/拉长后的形状。可以看出，这段语音信号具有很强的准周期性，且具有较大的振幅。它的周期即基音周期。图 c 是/K/辅音的展开图。可以看出，辅音的波形类似于白噪声，且振幅较小。



(a) 语音信号“开始”的确时域波形



(b) 元音部分/a/展开波形



(c) 辅音部分/k/的展开波形

图 4 语音信号“开始”的时域波形及其展开图

3.2.1.2 短时能量分析

语音信号的能量随时间变化比较明显，一般清音部分的能量比浊音的能量小得多。对语音信号的短时能量进行分析，可得出反映这些幅度变化的合适描述方法。对

于信号 $\{x(n)\}$ ，短时能量的定义如下：

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) = x^2(n) * h(n) \quad (3-3)$$

式 4-1 中 $h(n)=w^2(n)$ ， E_n 表示在信号的第 n 个点开始加窗函数时的短时能量。可以看出，短时能量可看作语音信号的平方经过一个线性滤波器^[2]的输出，该线性滤波器的单位冲激响应为 $h(n)$ ，如图所示：

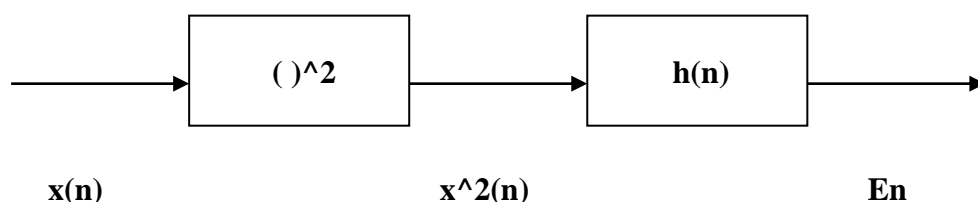


图 5 短时能量的方块图表示

冲激响应 $h(n)$ 的选择，或者说窗函数的选择决定了短时能量表示方法的特点。为了反映窗函数的选择对短时能量的影响，假设式(2-1)中 $h(n)$ 非常长，且为恒定幅度，那么 E_n 随时间的变化将变为很小，这样的窗函数等效为很窄的低通滤波器。我们需要对语音信号进行低通滤波，但太窄的低通滤波就不能反映语音信号幅度变化。因此出现了窗长选取上的矛盾，这种矛盾将在语音信号的短时表示方法的研究中反复出现。即希望有一个短时窗(冲激响应)，以响应快速的幅度变化。如果用 $x_w(n)$ 表示 $x(n)$ 经过加窗处理后的信号，窗函数的长度为 N ，则短时能量可表示为：

$$E_n = \sum_{m=n}^{n+N-1} x_w^2(m) \quad (3-4)$$

短时能量主要有以下几个方面的应用：首先利用短时能量可区分清音和浊音，因为 浊音的能量要比清音的能量要大得多；其次可以用短时能量对有声段和无声段进行判定，对声母和韵母分界，以及对连字分界等。

由于短时能量是对信号的平方运算，因而人为增加了高低信号之间的差距，在一些应用场合不大适用。解决这个问题的方法是采用短时平均幅度值来表示能量的变

化。

其公式为：

$$Mn = \sum_{-\infty}^{\infty} |x(m)| w(n-m) \quad (3-5)$$

3.2.1.3 短时平均过零率

短时间平均过零率是语音信号时域分析中最简单的一种特征。顾名思义，它是指每帧内信号通过零值的次数。对于连续的语音信号，可以考察其时域波形通过时间轴的情况。对于离散信号，短时平均过零率实质上就是信号采样点符号变化的次数。如果是正弦信号，它的平均过零率就是信号频率除以两倍的采样频率，而采样频率是固定的，因此，过零率可以在一定程度上反映其频谱特性，可以通过短是过零率获得频谱特性的一中粗略估计。短时平均过零率的公式为：

$$Zn = 1/2 \sum_{-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (3-6)$$

其中：

$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \quad (3-7)$$

短时平均过零率可以用于语音信号分析。在发浊音的时候声带振动，因而声门激励是频率为基频的声压波，它在经过声道时产生共振。尽管声道有若干个共振峰，但由于声门的影响，其能量分布主要集中在 3KHz 频率范围内；反之，在清音段声带不振动。因此，浊音段时的能量集中在低频段，而清音的能量集中的高频段。由于短时平均过零率可以在一定程度上反映频率的高低，因此在浊音段一般有较低的过零率，而在清音段具有较高的过零率，这样就可以用短时平均过零率来初步判断清音和浊音。

3.2.2 语音信号的频域分析

时域波形虽然简单直观,但对于语音这样复杂的信号而言,一些特性要在频域中才能体现出来;并且无论是从发音器官的共振角度,还是从听觉器官的频率响应角度来看,频谱都是表征语音特性的基本参数。其中共振峰就是一个典型的频域参数,它可以决定信号频谱的总体轮廓或包络。对于声道而言,它的共振峰频率不止一个,一般元音可以有 3~5 个共振峰。

在语音的发音过程中,声道通常都处于运动状态,这个运动状态的时变过程同振动过程相比要缓慢得多,因此一般假设语音信号是一种短时平稳信号,在一个很短的时间内(10ms-30ms)是相对平稳的,但在长时间的周期中语音信号的特性会发生变化,这种变化的不同决定了产生语音的不同。根据语音信号的这种短时平稳的特点,在每一时刻都可以用该附近的短语音信号分析得到一个频谱。图 6 中给出了“开始”中/a/的频谱特性。其中横轴表示频率,纵轴表示该频率的强弱,以分贝(dB)为单位。这里短时分析采用汉明窗,进行频谱分析的窗长为 300 个采样点。

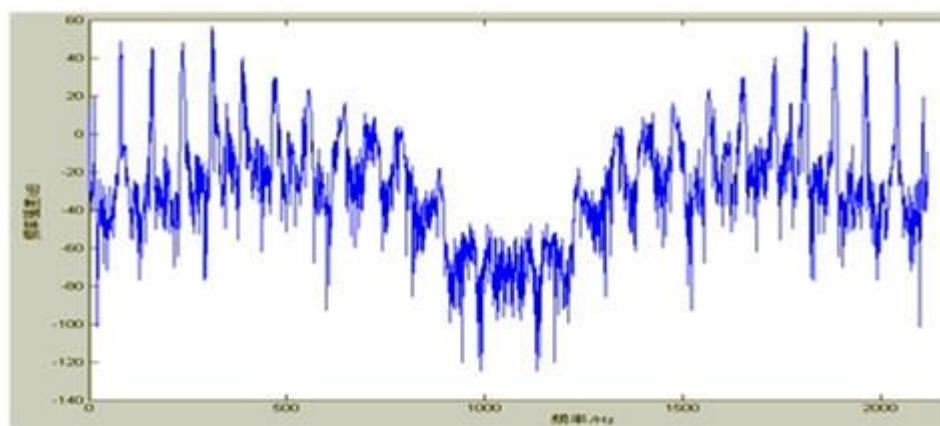


图 6 “开始”中/a/的频谱特性

语音的感知过程与人类听觉系统具有频谱分析功能是紧密相关的。因此,频谱分析是认识语音信号和处理语音信号的重要方法。在频域上研究语音信号,可以使某些在时域上无法体现的特性变得十分明显。

常用的频域分析方法有带通滤波器组法、傅立叶变换法、和线性预测法等,傅立叶分析是分析线性系统和平稳信号稳态特性的强有力手段,傅立叶变换可以将信号分解为各个不同频率分量的组合,从而把信号的时域特征和频域特征联系起来。

但是，傅立叶变换使用的是一种全局变换，无法标明信号的时频局域性质。为了能够分析和处理非平稳信号，人们对傅立叶变换进行改进提出了倒谱法，倒谱法是将对数功率谱进行反傅立叶变化后得到的，它可以进一步将声道特性和激励特性有效地分开，因此可以更好的揭示语音信号的本质特征。本文将会重点介绍有关倒谱域的基音检测算法。

3.3 几种典型的基音检测算法分析

准确地检测语音信号的基音周期，对于高质量的语音分析与合成、语音压缩编码、语音识别和说话者确认等方面具有重要的意义^[2]。从前言部分可知，由于影响基音检测的因素很多。因此，很难从复杂多样的语音信号中提取仅和声门振动有关的音源信息。然而尽管困难很大，但因为它的重要性，基音的检测一直是很值得研究的课题。

本章节将对自相关函数法、平均幅度差函数法、小波变换法、倒谱法这四种传统的基音周期估计方法的原理进行分析。

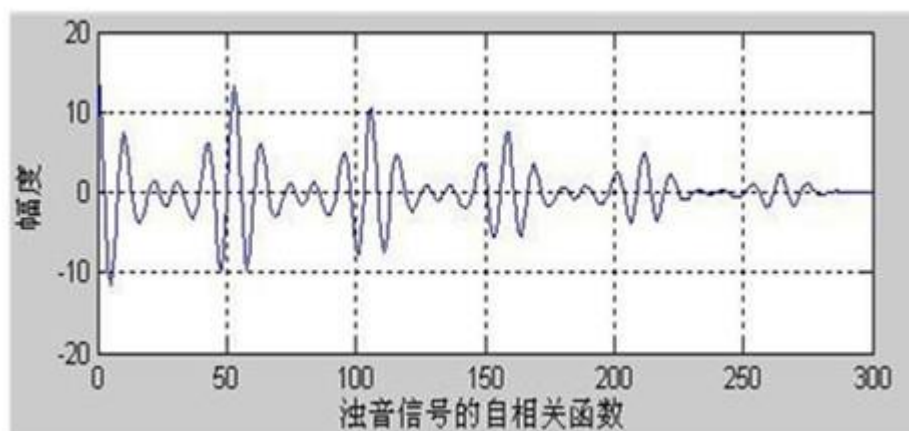
3.3.1 自相关函数法

浊音信号的短时自相关函数在基音周期的整数倍位置存在较大的峰值，而清音的短时自相关函数没有明显的峰值出现。因此检测自相关函数是否有峰值，就可以判断是清音还是浊音，且两相邻峰值之间对应的时间段就是基音周期。

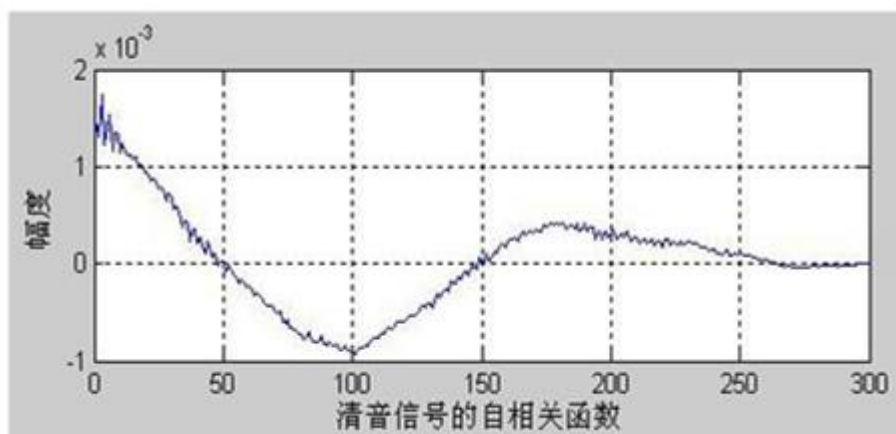
设 $s_w(n)$ 是一段加窗语音信号，它的非零区间为 $n=0 \sim (N-1)$ 。 $s_w(n)$ 的自相关函数称为语音信号 $s(n)$ 的短时自相关函数，用 $R_w(k)$ 表示，它的计算分式为：

$$R_w(k) = \sum_{n=-\infty}^{\infty} s_w(n)s_w(n+k) = \sum_{n=0}^{N-k-1} s_w(n)s_w(n+k) \quad (3-8)$$

由短时自相关函数的定义可以看出其所具有的一些性质：① $k=0$ 时函数取得最大值，此时，自相关函数的取值就是该信号的短时能量；② 如果原序列是周期为 T 的周期信号，那么自相关函数也是周期为 T 的周期函数。



(1) 浊音信号自相关



(2) 清音信号自相关

图7 清音段与浊音段自相关函数比较

从图中找出第一最大峰值的位置就可以估计出基音周期。但有时第一最大峰值位置并不一定与基音周期吻合。基音周期估计与声道特性的影响有关，主要是由于声道的共振峰特性会对基音周期估计造成干扰。这是因为语音信号包含丰富的谐波分量。基音频率的范围分布在 50-450Hz 左右，其中 100-250Hz 的情况占大多数，所以浊音信号有可能包含 30-40 个谐波分量。同时，由声道特性决定的语音信号的第一共振峰通常在 300-1000Hz 的范围内，这样就有可能导致语音的第 2-8 个谐波分量的幅度高于基频分量。因此，丰富的谐波分量常常会导致基音周期估计出现“倍频”或“半频”错误。

为了减少共振峰的影响,需要进行一些前期的预处理。其中中心削波法是一种有效的非线性处理方法,它可以消除语音信号的低幅值部分,数学表达式为:

$$y(n) = C(x(n)) = \begin{cases} x(n)-L, & x(n) > L \\ x(n)+L, & x(n) < -L \\ 0, & |x(n)| \leq L \end{cases} \quad (3-9)$$

式中,削波电平 L 由语音信号的峰值幅度来确定,它等于语音段最大幅度的一个固定百分数。一般取最大信号幅度 60%~70%。这个门限的选择是重要的,一般在不损失基音信息的情况下应尽可能选得高些,以便达到较好的效果。经过中心削波后只保留超过削波电平的部分,其结果是削去了许多和声道响应有关的波动。对中心削波后的语音再计算自相关函数,这样在基音周期位置呈现大而尖的峰值,而其余的次数的峰值幅度都很小。

其削波特性如下:

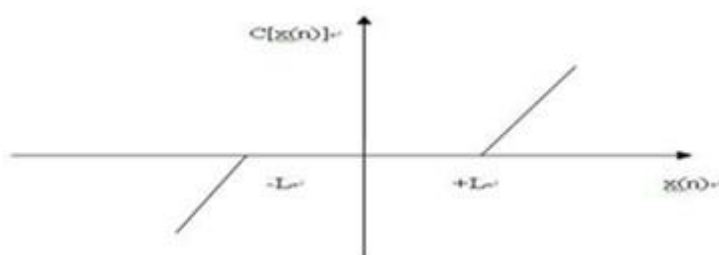


图 8 中心削波函数

下图画出了未削波与削波后计算自相关函数的波形。

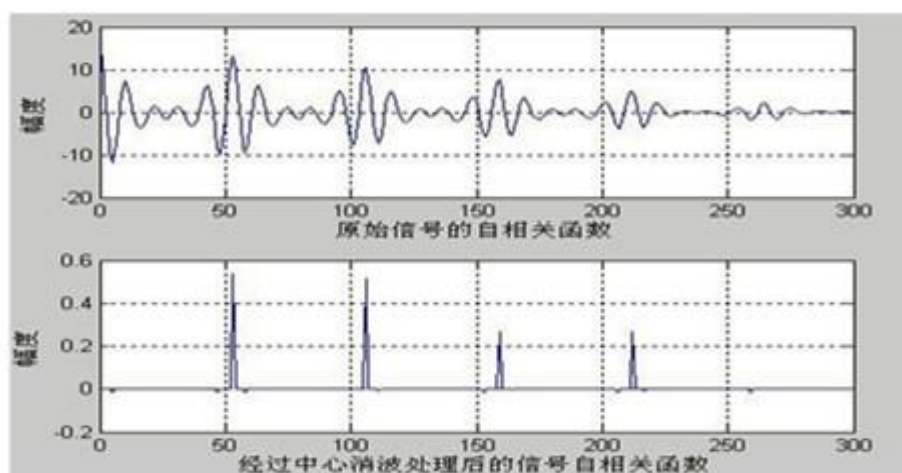


图 9 原始信号与中心消波后信号自相关函数比较

此外，还可以用一个通带为 900 Hz 的线性相位低通滤波器除去高次谐波分量。这样处理后的信号基本就只含有第一共振峰一下的基波和谐波分量。实验表明，用这种方法做预处理，对改善自相关和平均幅度差函数的基音检测都有明显的效果。如图所示原始信号的自相关函数和经过 900Hz 低通滤波后的波形图。

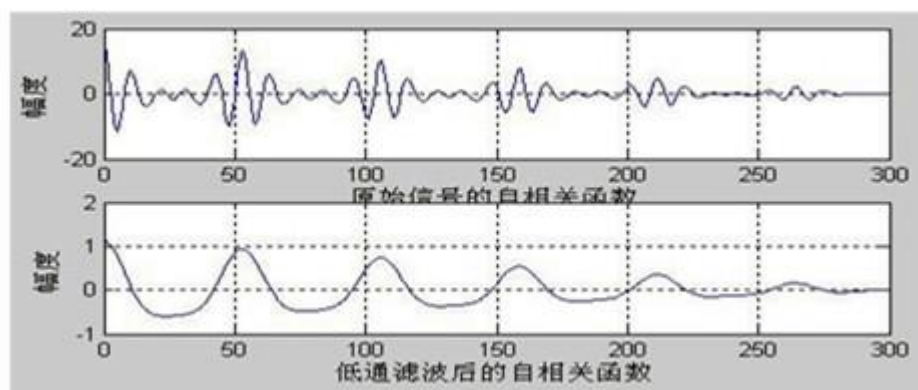


图 10 原始信号与低通滤波后信号自相关函数比较

3.3.2 平均幅度差函数法(AMDF)

短时自相关函数是语音信号时域分析的重要参数。但计算自相关函数的运算量很大，其原因是乘法运算所需要的时间较长。简化计算自相关函数的方法有很多，但都

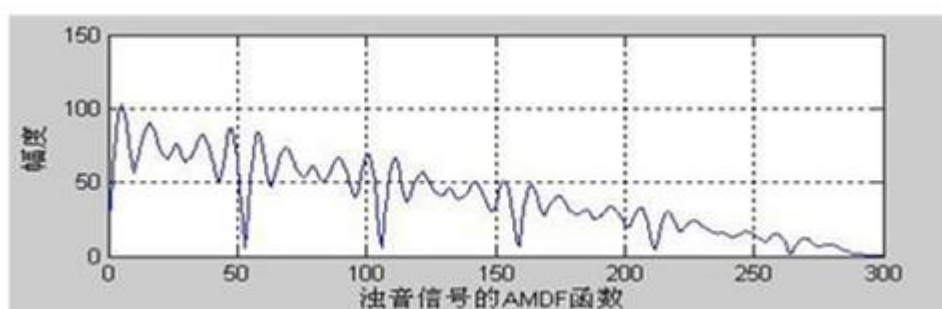
无法避免乘法运算。为了避免乘法运算,常常采用另一种与自相关有类似作用的参量,即短时平均幅度差函数(AMDF)。它是基于这样一个想法: 对一个单纯的周期信号求差值, 即:

$$d(n) = s(n) - s(n-k) \quad (k=0, \pm T_p, \pm 2T_p, \dots) \quad (3-10)$$

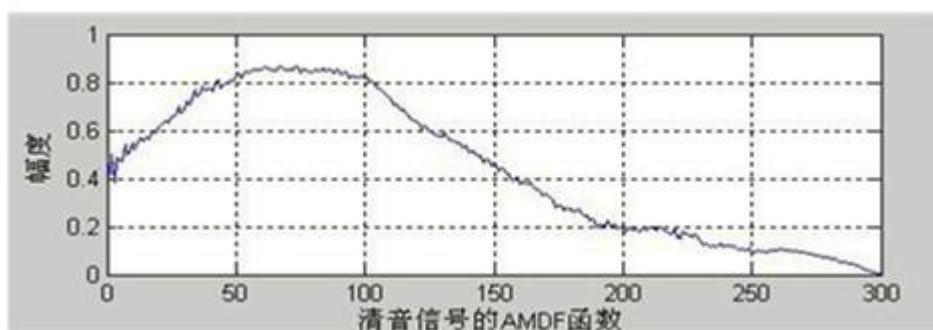
则在 $k=0, \pm T_p, \pm 2T_p, \dots$ 时, 上式为零。即当 k 与信号周期吻合时, 作为 $d(n)$ 的短时平均幅度值总是很小, 因此短时平均幅度差函数的定义为:

$$Fn(k) = \sum_{n=0}^{N-k-1} |Sw(n+k) - Sw(n)| \quad (3-11)$$

由上式可知, 对于周期性的 $d(n)$, $F_n(k)$ 也呈现周期性。与短时自相关函数不同的是在周期的整数倍点上 $F_n(k)$ 具有的是谷值, 而非峰值。因此在浊音信号的语音周期内, $F_n(k)$ 值迅速下降, 而在清音语音时, 不会有明显的下降。由此可见, 短时平均幅度差函数可以用于基音周期的检测。而且计算上比短时自相关函数方法更为简单。下图分别为浊音和清音段的 AMDF 函数:



(1) 浊音段的 AMDF 函数



(2) 清音段的 AMDF 函数

图 11 清音段与浊音段 AMDF 函数比较

由于计算 AMDF 函数只需要加、减和取绝对值运算，当用专用硬件或定点 DSP 系统实现时，与自相关函数相比，运算量大大减少。同时，AMDF 函数在基音周期点的谷值比自相关函数的峰值更加尖锐，因此错判率相对较小，稳分建性更高。但当语音信号的幅度快速变化时，AMDF 函数的谷值深度会减小，从而影响基音估计的精度。

3.3.3 小波变换法

小波变换是 20 世纪 80 年代中后期逐渐发展起来的一种数学分析方法。所谓小波就是小的波形。“小”指其衰减性，“波”指其波动性，小波理论采用多辩率分析的思想，非均匀地划分时频空间，这使得小波变换在非平稳信号的分析 and 处理中发挥着重要的作用。由于语音信号是一种比较典型的非平稳信号，因此很多学者将小波变换引入到语音信号处理中，并展开了相关的研究工作。下面将主要介绍其在基音检测方面的应用。

如果一个函数 $\theta(x) = o(1/(1+x^2))$ ，且它的积分非零，则称 $\theta(x)$ 为平滑函数。一般地，平滑函数的能量通常集中在低频段，所以，平滑函数可看成低通滤波器的系统响应函数。由于这个原因， $f(x)$ 与 $\theta(x)$ 的卷积 $(f * \theta)(x)$ 衰减了 $f(x)$ 的高频信息，而不改变低频部分，从而平滑了 $f(x)$ 。

由小波变换与信号奇异性的关系可知，如果选择小波函数为平滑函数的一阶导数，则小波变换 $Wf(s, x)$ 的模极大值点对应于 $f(x)$ 的突变点(奇异点)。可以利用小波变换的模极大值点来检测出语音信号的突变点，也就是检测也声门闭合的时刻，基音周期就是两次突变的时间间隔。

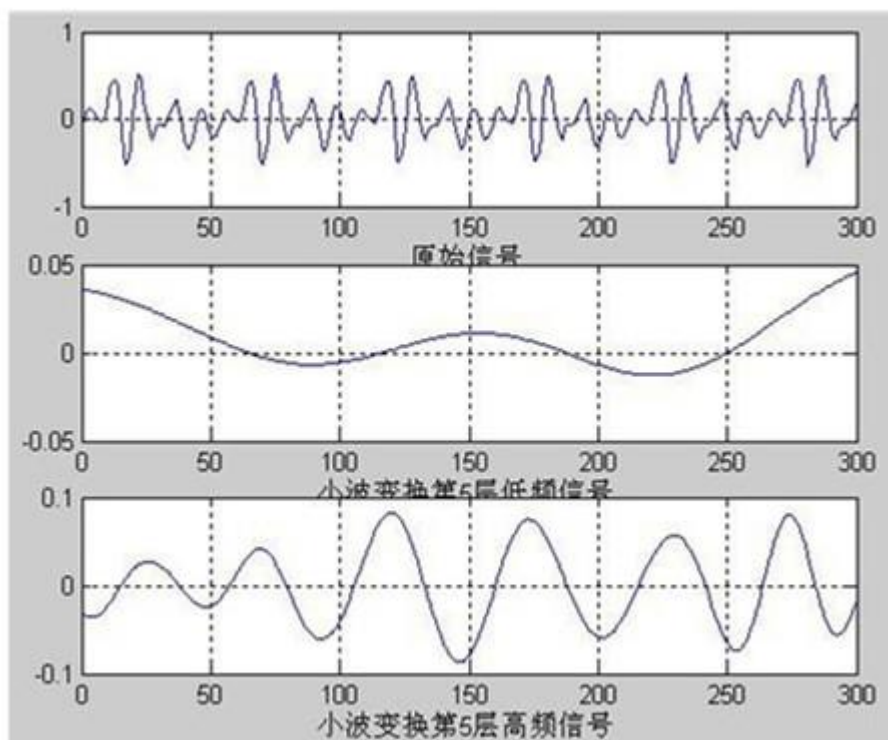


图 12 浊音段小波变换波形法进行基音检测

根据小波变换的最大级数确定公式 $I = \log_2(f_s/f_0)$ ，其中 f_s 为信号的采样频率， f_0 为欲在信号中保留的最高频率。在 f_s 为 8000Hz， f_0 取 500Hz 时， I 取值为 4，即在对语音信号进行小波变换时，变换级数至少可以达到 4 级，其逼近部分仍保留有基音频率信息。大多数的语音样本，在采用小波变换阶数为 5 或 6，连续进行五级变换，一般可以得到满意的分析效果。

小波变换法不是建立在语音信号短时平稳性的基础上的，所以它能提取出精确反映基音周期变化的动态基频包络，其他的短时分析的基音检测方法，由于都是基于短时平稳假设的基础上的，所以实际上求出的是基音周期在某一段时间的平均值。而小波变换法恰恰改正了这一固有的弱点，它能动态地随语音信号的周期性变化而变化，不会因为语音的准周期性或某一段无周期性而影响提取效果。

所以，利用小波变换进行基音检测，可以随说话人的不同、语音段的不同精确地检测出基音周期，从而构成了真正反映基音周期变化的基频包络。但当检测被噪声污染的语音信号时，随着信噪比的减小，利用小波变换进行基音检测的误差越来越大。

3.3.4 倒谱法

在介绍的语音生成机理时，曾提到语音信号的生成模型可以看作是声门的激励信号和声道的冲激响应的卷积。在语音信号处理的实际应用中，很多场合需要根据语音信号反过来求解声门信号或声道冲激响应。这就需要在知道卷积结果的基础上，利用“解卷”求得参与卷积的各个信号，同态处理是常用的解卷方法。由于语音信号进行同态分析后得到的是语音信号的倒谱参数，因此同态分析也叫倒谱分析^[21]。

Z 变换可以将卷积信号转变为乘积信号，而取对数运算可以将乘法转化为加法，卷积同态处理正是采用这样的一种思路，具体过程如下：

① 输入信号：

$$x(n) = x_1(n) * x_2(n) \quad (3-12)$$

② Z 变换：

$$X(z) = Z[x(n)] = Z[x_1(n)] Z[x_2(n)] = X_1(z) X_2(z) \quad (3-13)$$

③ 取对数运算：

$$\ln X(z) = \ln X_1(z) + \ln X_2(z) = X^{\wedge}(z) = X_1^{\wedge}(z) + X_2^{\wedge}(z) \quad (3-14)$$

④ 逆 Z 变换：

$$x^{\wedge}(z) = Z^{-1}[X^{\wedge}(z)] = Z^{-1}[X_1^{\wedge}(z)] + Z^{-1}[X_2^{\wedge}(z)] = x_1^{\wedge}(n) + x_2^{\wedge}(n) \quad (3-15)$$

⑤ $x(n)$ 为加性信号，经过线性系统的处理得到：

$$y^{\wedge}(n) = y_1^{\wedge}(n) + y_2^{\wedge}(n) \quad (3-16)$$

⑥ 对 $y(n)$ 作 Z 变换：

$$Y^{\wedge}(z) = Z[y^{\wedge}(n)] = Z[y_1^{\wedge}(n)] Z[y_2^{\wedge}(n)] = Y_1(z) Y_2(z) \quad (3-17)$$

⑦ 指数运算：

$$Y^{\wedge}(z) = \exp[y^{\wedge}(n)] = \exp[y_1^{\wedge}(n)] \exp[y_2^{\wedge}(n)] = Y_1(z) Y_2(z) \quad (3-18)$$

⑧ 逆 Z 变换：

$$y(n) = Z^{-1}[Y(z)] = Z^{-1}[Y_1(z)] Z^{-1}[Y_2(z)] = y_1(n) * y_2(n) \quad (3-19)$$

步骤①～④称作特征系统，记为 $Z[\]$ ；⑥～⑧步骤称为逆特征系统，记为 $Z^{-1}[\]$ 。我们称中的时域序列 $x(n)$ 为信号序列 $x(n)$ 的“复倒频谱”，简称“复倒谱”，也叫做对数复倒谱。即：

$$x^{\wedge}(n) = Z^{-1}[\ln[x(n)]] , \quad (3-20)$$

$x(n)$ 所在的时域被称作复倒谱域。又注意到，有

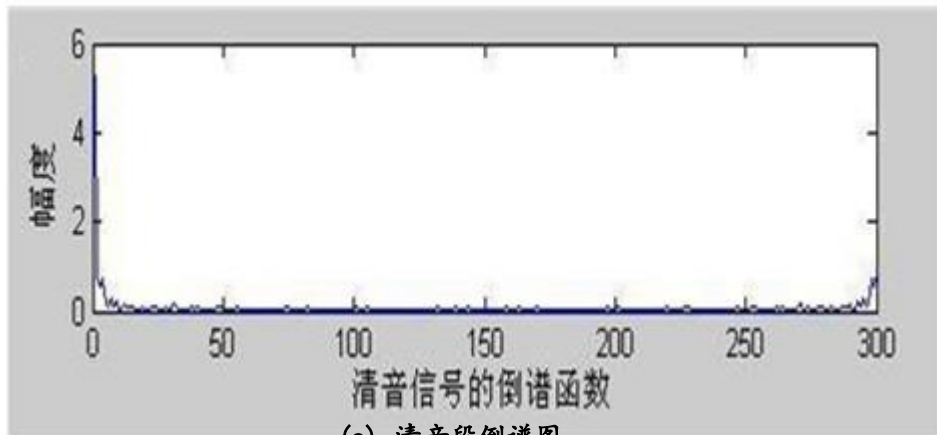
$$\hat{X}(z) = \ln(X(z)) = \ln(|X(z)|e^{-j\theta}) = \ln(|X(z)|) - j\theta, \quad \theta = \arg X(z) \quad (3-21)$$

如果仅对 $\hat{X}(z)$ 实部作逆 Z 变换: $c(n) = Z[\ln|X(z)|]$

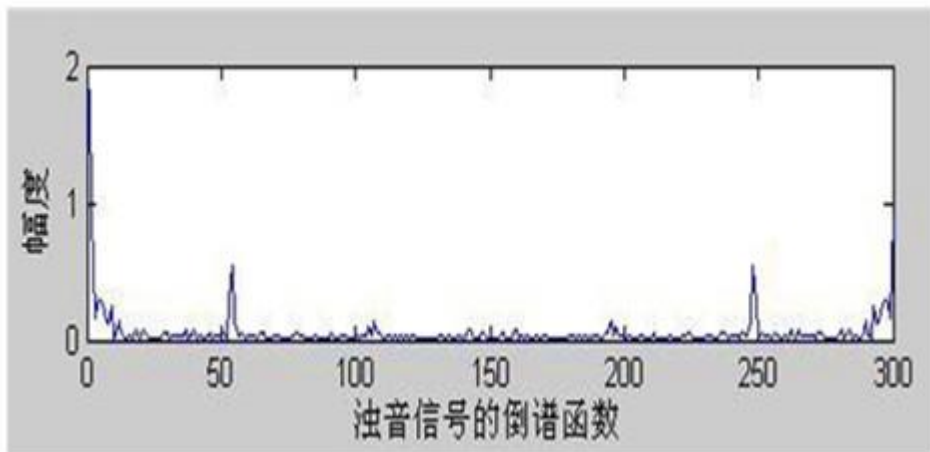
称 $c(n)$ 为“倒频谱”，简称“倒谱”，也称“倒频”。值得注意的是，倒谱不能通过逆特征系统还原成自身，因为在计算中相位信息丢失了。

在绝大多数的应用场合，特征系统和逆特征系统中的正反 Z 变换都可以用正反离散傅里叶变换来代替，此时倒谱的定义变为

$$c(n) = \text{IDFT}\{\ln|\text{DFT}[s(n)]|\} \quad (3-22)$$



(a) 清音段倒谱图



(b) 浊音段倒谱图

图 13 清音段与浊音段倒谱图比较

在频域内，语音信号的短时谱等于激励源频谱与滤波器频谱的乘积，浊音信号的短时谱中包含的快变化的周期性细致结构，必定对应着周期性脉冲激励的基频和各次谐波。语音的倒谱是将语音的短时谱取对数后再进行 IDFT 得到的，所以浊音信号的周期性激励反映在倒谱上是同样周期的冲激。借此，可从倒谱波形中估计出基音周期。一般把倒谱波形中第二个冲激，认为是对应激励源的基频，即基音周期。

4 过程论述

4.1 基于 MATLAB 的基音测检方法的实现与分析

4.1.1 MATLAB 软件简介

MATLAB 是 Math Works 公司于 1982 年推出的一套高性能的数值计算和可视化软件，它集数值分析、矩阵运算、信号处理和图形显示于一体，构成了一个方便的、界面友好的用户环境。其强大的扩展功能为各个领域的应用提供了基础。主要涉及信号处理，控制系统、神经网络、图像处理、鲁棒性控制、非线性系统控制设计、系统辨识等工具箱，这些工具箱给各个领域的研究和工程提供了有力的工具。借助这些工具，各个层次的研究人员可直观、方便地进行分析、计算及设计工作，从而大大地节省了时间。

4.1.2 自相关函数法进行基音检测

本实验的目的是找出一种能运用自相关函数法准确检测出一帧语音信号基音周期的算法，并测试此算法的抗噪能力。

4.1.2.1 算法分析与设计

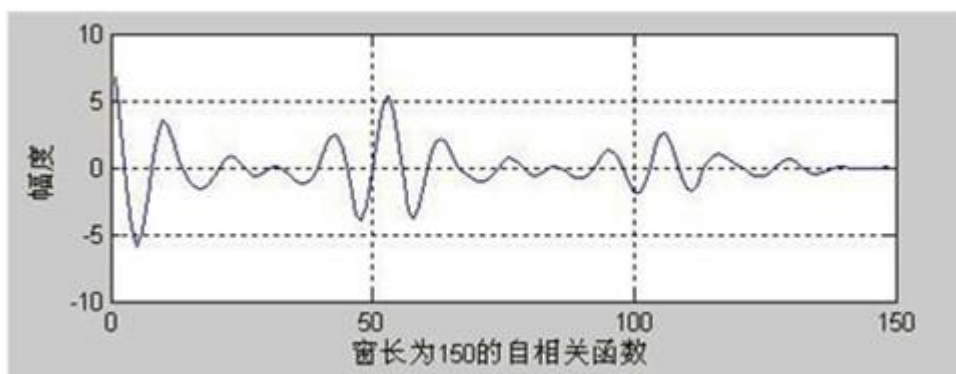
该算法选择汉明窗，因为汉明窗是窗函数的选取原则为窗函数截取后的 $x(n)$ 尽量是中间大两头小的光滑函数，冲激响应对应的滤波器具有低通特性，能使短时平均能量 E_n 能反映语音信号的幅度变化。

本文所有实验均以采样频率为 10KHz、精度为 16 比特的 wav 文件作为声源。因为语音信号包含非常丰富的谐波分量，基音频率最低可达 80Hz，最高可达 900Hz，但基音频率大多数分布在 100-200Hz 之间。因此，语音信号可能含有三四十次谐波分量，而其基波分量往往不是最强的分量。语音信号的第一共振峰通常在 300-1000Hz 范围内，即基音的 2-8 次谐波成分比基波分量还要强。为了提高检测的准确度，实验中引入了一个 900Hz 低通滤波模块滤除语音帧的高次谐波分量。

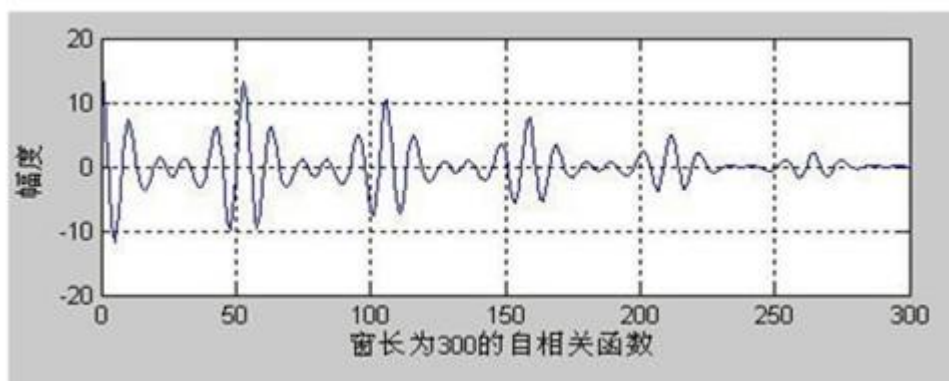


图 14 短时自相关函数基音检测示意图

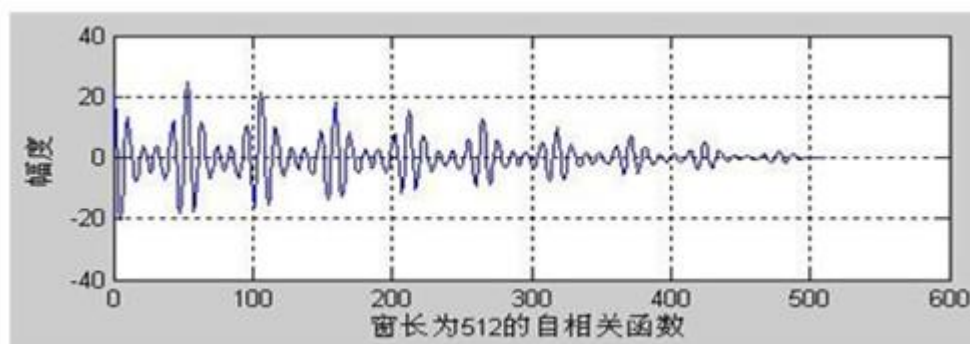
实验中采用的语音信号经短时能量分析找出浊音段后，取 500~799 采样点共 300 个(一帧信号)采样点进行分析。因为一帧语音信号有 300 个采样点，所以短时加窗处理时的窗函数的长度也是 300。下面先来说明本实验为什么取 300 个采样点为一帧语音信号，即，采用长度为 300 的窗函数短时处理的原因。



(a) 窗长为 150 的自相关函数



(b) 窗长为 300 的自相关函数



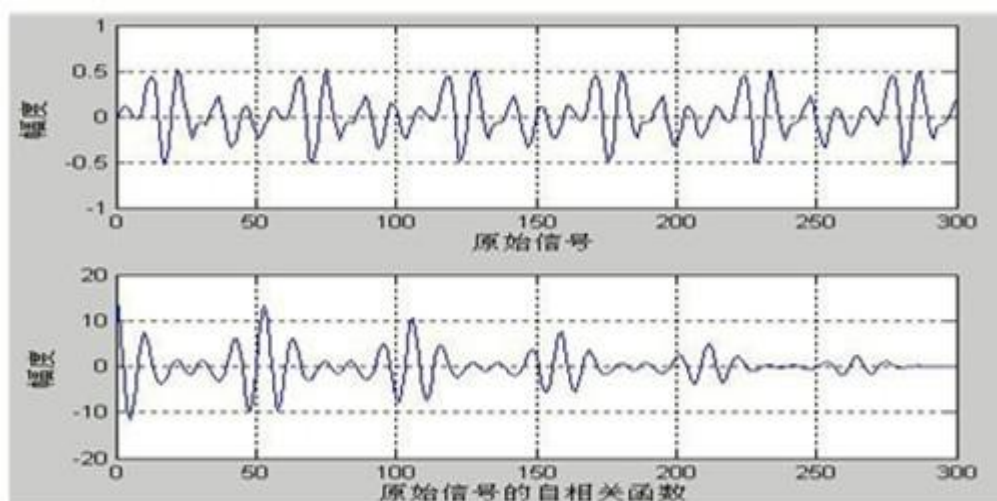
(c) 窗长为 512 的自相关函数

图 15 窗函数 N 值的不同对基音周期检测的影响

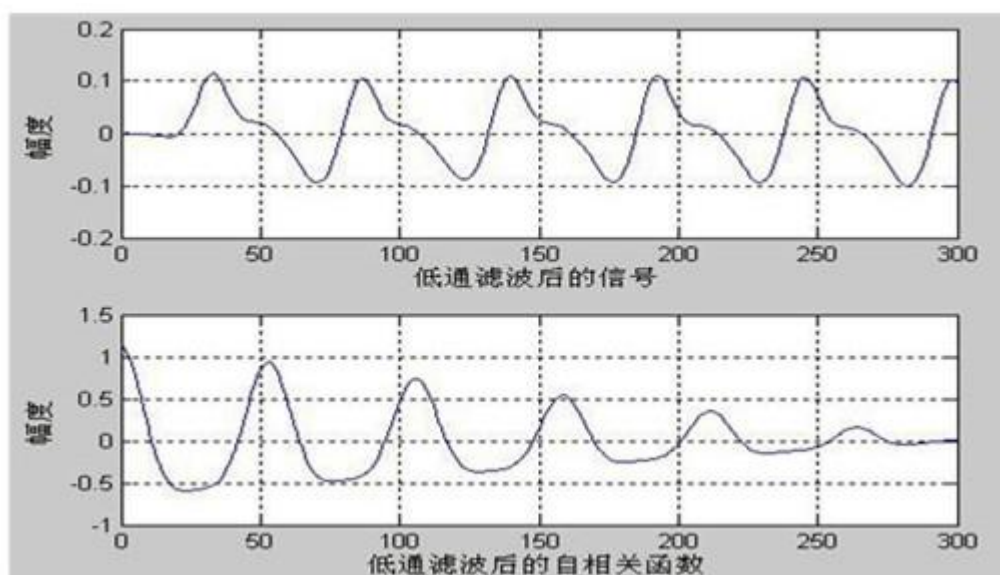
上图采样频率为 10kHz 时，一帧语音信号分别选取 150、300、512 样点时短时自相关函数的波形图。由图可知，窗函数长度太长时(图 c)，波形的细节部分不够清晰，基音周期检测结果不准确；窗函数的长度太短时(图 a)，一帧信号只有一个或者不到一个基音周期，波形变化太缓慢，也不利于检测基音周期；只有窗函数的长度选择适当(图 b)，一帧内含有 4 个基音周期，波形变化明显，峰值点尖锐，能够准确读取基音周期。

4.1.2.2 实验过程及仿真结果

本次实验用短时自相关函数法共做了四次实验仿真。第一次是用原始信号不加任何处理进行自相关分析；第二次把语音信号经过一个 900Hz 的低通滤波器后，进行自相关分析；第三次是把语音信号经过中心消波处理后，进行自相关分析；最后是把原始信号加强度分别为 0.05、0.25、0.5 噪声后，再短时自相关分析。实验结果如下：

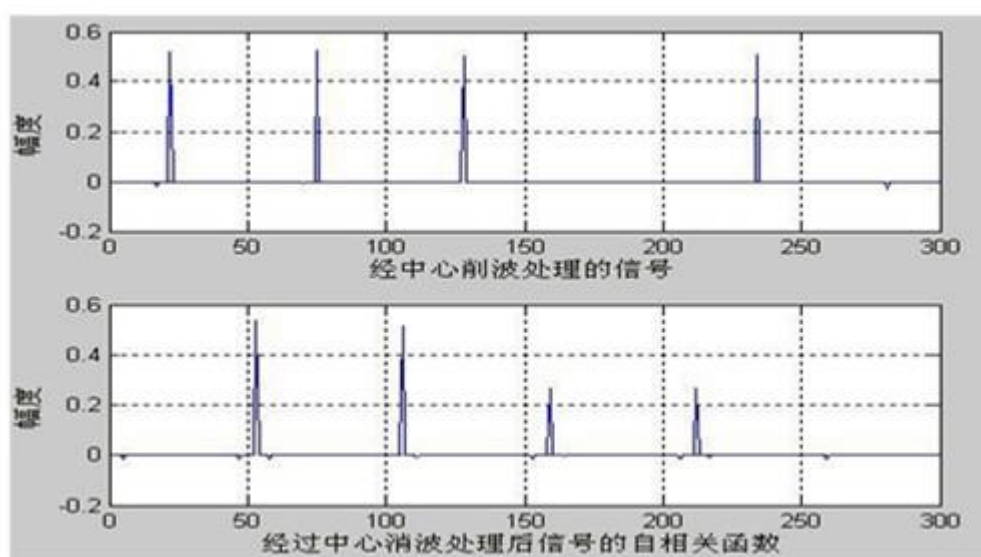


(1) 原始信号的自相关分析



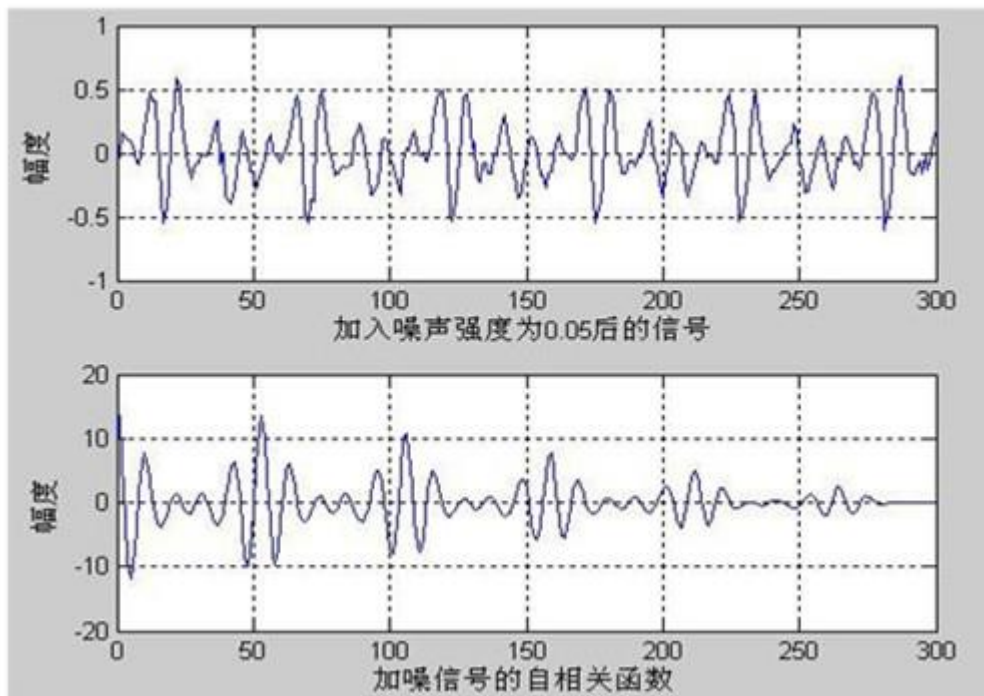
(2) 低通滤波后的自相关分析

对比图(1)和图(2)可以看出原始信号波形不光滑，含有较多的高频分量其自相关函数波形也不够光滑，经低通滤波后变得光滑，波形没有较多的突变，但峰值仍然不够尖锐。

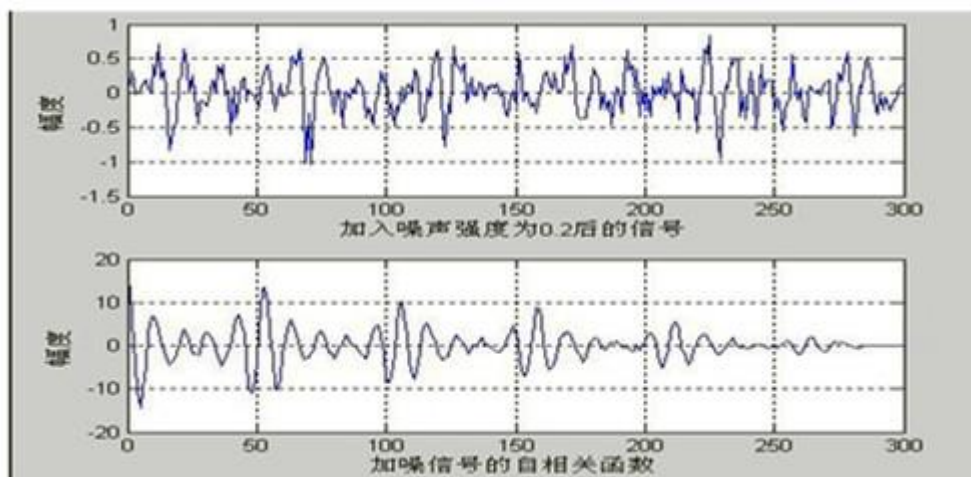


(3) 中心削波后的自相关分析

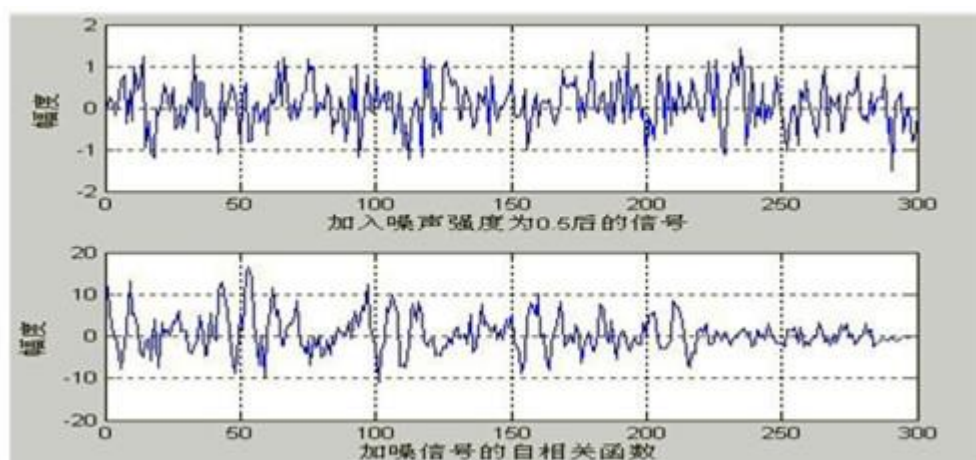
由图(3)可以看出经中心消波处理后，只有原始语音信号幅值高于特定值时才有值，低于特定值的信号全部为零；其自相关相函数波形在特定区域有取值，峰值点非常尖锐。



(4) 加噪(强度为 0.05)后的自相关分析



(5) 加噪(强度为 0.2)后的自相关分析



(6) 加噪(强度为 0.5)后的自相关分析

图 16 实验仿真结果

随着噪声强度的加强，函数波形越来越模糊，以致不能分辨最值点。根据仿真结果图估算出基音周期：

从图 16(1)可以看出，第一个峰值的位置出现在约 60 的滞后点上。因为浊音的自相关函数具有一定的周期性，在相隔一定的取样后，自相关函数达到最大值。浊音语音的周期可用自相关函数中的第一个峰值的位置来估算。

此段浊音的基音周期计算为： $T_1 = (0.1 \times 60) / 1000 = 6.0\text{ms}$ ，

同理可以估算出： $T_2 = 6.3\text{ms}$ ， $T_3 = 6.5\text{ms}$ ，

不同的噪声环境下估算的基音周期分别为 $T_4 = 7.0\text{ms}$ ， $T_5 = 7.8\text{ms}$ ，

T_6 无法从图中估算出。

由此可以看出，用自相关函数法来进行基音周期的检测，简单直观、运算量不大，但是在没有进行预处理(低通滤波或中心消波)时的波形受 50Hz 电源和共振峰影响，波形不光滑，不利于基音周期的估计；经过预处理波形变得尖锐，能较准确基音估计，但是抗噪性能不强。

4.1.3 倒谱法进行基音检测

4.1.3.1 算法分析与设计

对语音信号解卷积原理，可以得出激励序列的倒谱，它具有与基音周期相同的周

期，因此可以容易精确地求出基音周期，图 6-4 为语音信号对数频谱示意图，它包含两个分量：对应与频谱包络的慢分量以及对应与基音谐波峰值的快分量。通过滤波或再取一次傅立叶反变换，即可将慢分量与快分量分开。图 6-5 为倒谱示意图，其中靠近原点的低频部分是频谱包络的变换，而位于窄峰处的则为谐波峰值的变换，表示基音。基音峰值的变换与频谱包络变换之间的间隔总是足够大，因而能很容易地对前者进行识别。

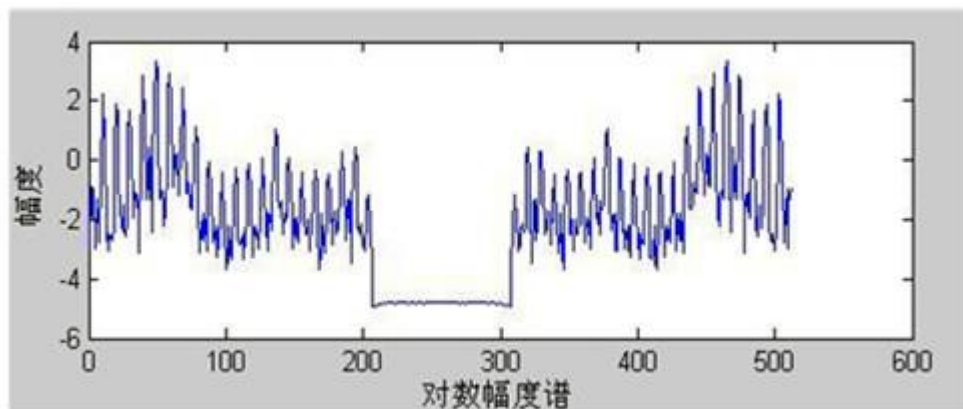


图 17 对数幅度谱

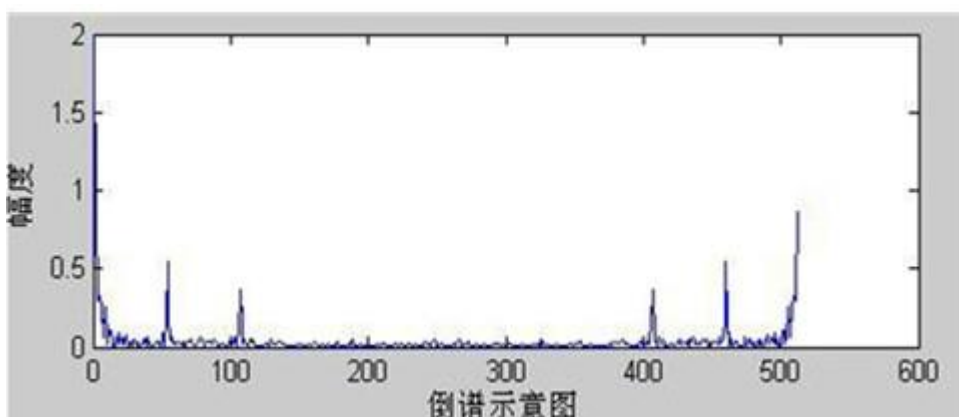


图 18 倒谱示意图

关于倒谱法的基本原理已经详细介绍过，这里就不再重复了，下面就用倒谱法来对一段语音进行周期性的检测。其工作原理简要说明如下：

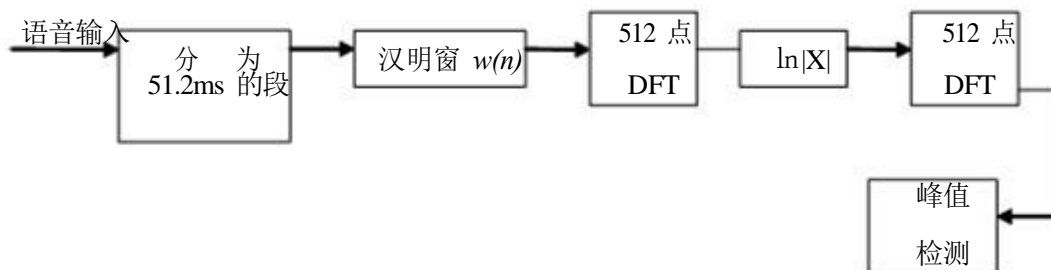
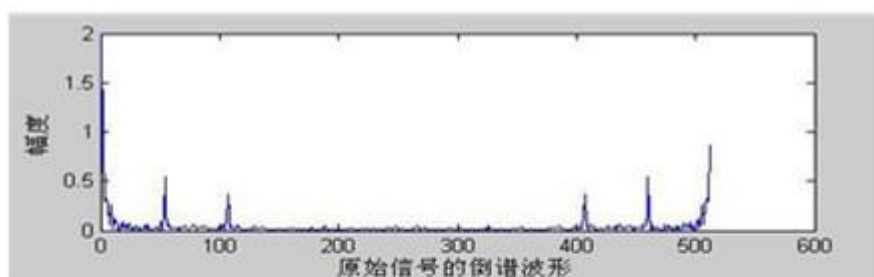


图 19 基音检测的倒谱法

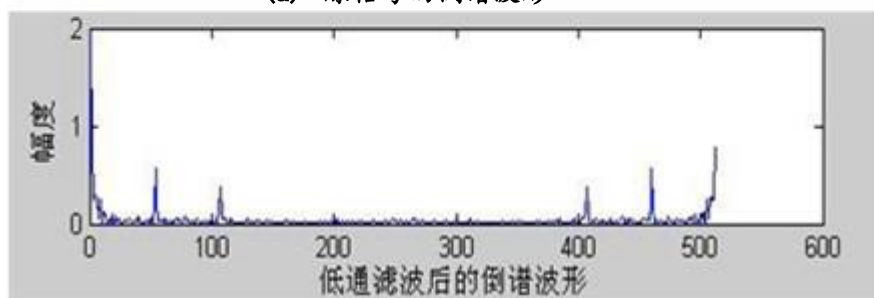
采样频率为 10kHz, 帧长为 51.2ms, 用汉明窗平滑, 然后经 512 点的傅立叶变换, 求出变换后的对数, 再进行一次傅立叶的反变换, 最后求出倒谱峰值的位置, 即基音周期。汉明窗的长度以及窗相对于语音信号的位置, 对倒谱峰的高度有相当大的影响, 为使倒谱具有明显的周期性, 窗口选择的语音段至少应包含两个明显的周期。

4.1.3.2 实验过程及仿真结果

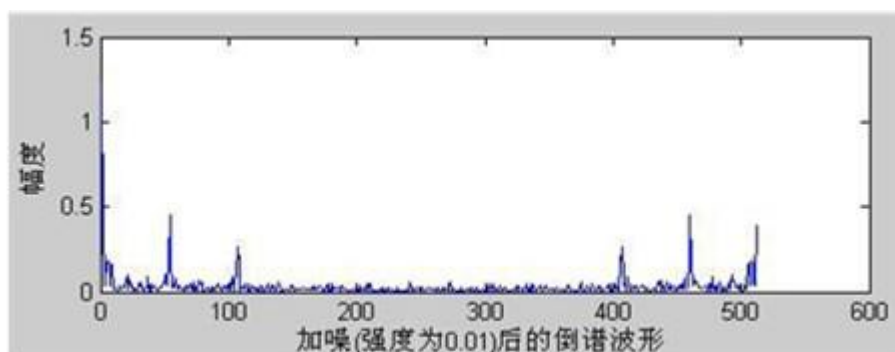
本次实验共做了三次实验仿真。第一次是用原始信号不加任何处理进行倒谱分析; 第二次把语音信号经过预处理即通过一个 900Hz 的低通滤波器后, 进行倒谱分析; 最后是把原始信号加强度分别为 0.01、0.03、0.1 噪声后, 倒谱分析。实验结果如下:



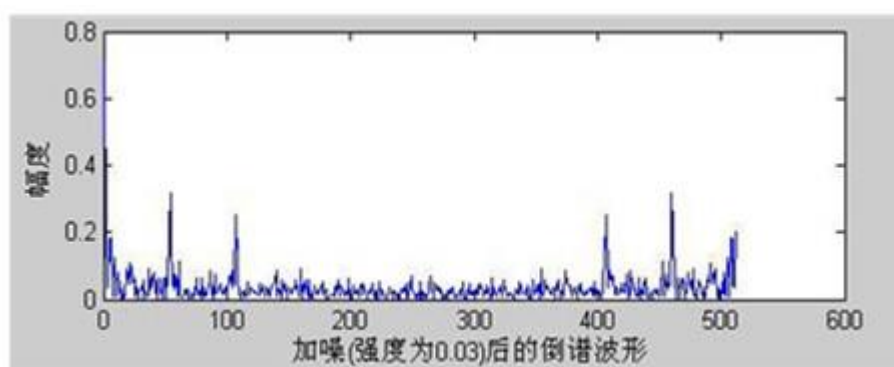
(a) 原信号的倒谱波形



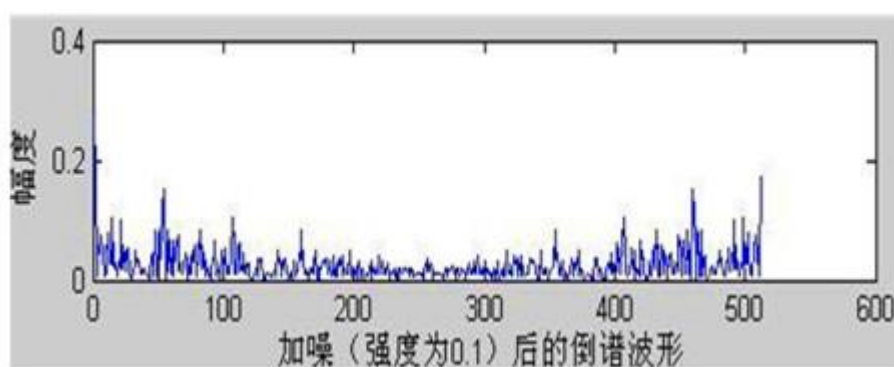
(b) 低通滤波后的倒谱波形



(c) 加噪（强度为 0.01）后的倒谱波形



(d) 加噪（强度为 0.03）后的倒谱波形



(e) 加噪（强度为 0.1）后的倒谱波形

图 20 三种不同方法的倒谱波形

此段浊音的基音周期计算为： $T_1 = (0.1 \times 58) / 1000 = 5.8\text{ms}$ ，

同理可以估算出 $T_2 = 6.6\text{ms}$ ， $T_3 = 7.8\text{ms}$ ，

不同的噪声环境下估算的基音周期分别为 $T_4=8.2\text{ms}$,

T_5 无法从图中估算出。

可以看出,采用无噪声语音时,用倒谱法进行基音检测是很理想的。然而当存在加性噪声时,在对数功率谱中的低电平部分被噪声填满,掩盖了基音谐波的周期性。这意味着倒谱的输入不再是纯净的周期性成分,而倒谱中的基音峰值将会变宽,并受到噪声的污染,从而导致倒谱检测方法的灵敏度随之下降。为此,本文提出了如下改进方法:

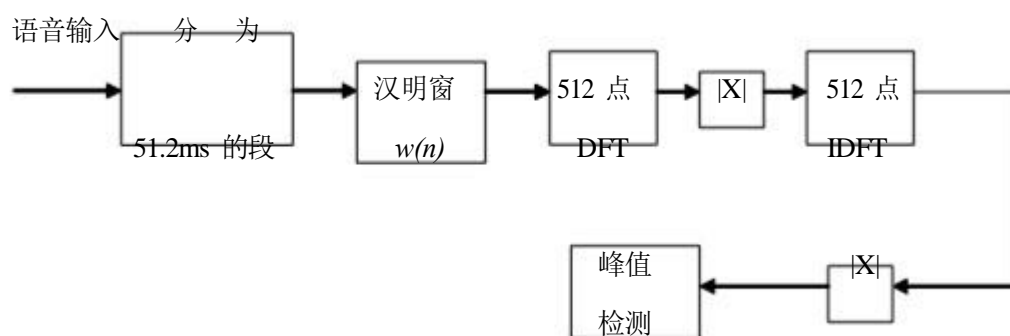
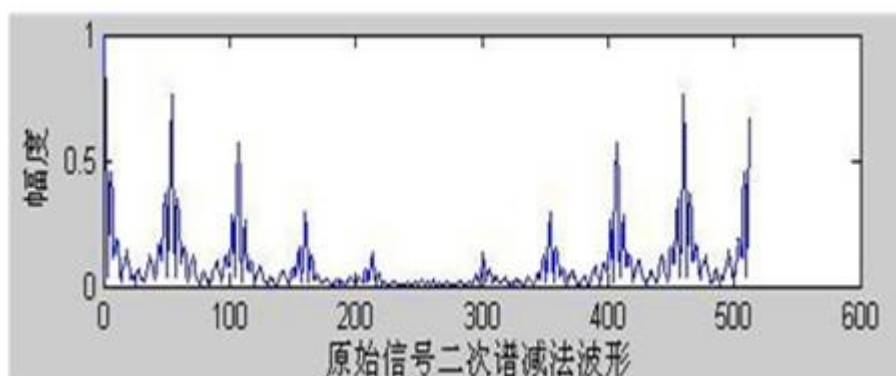


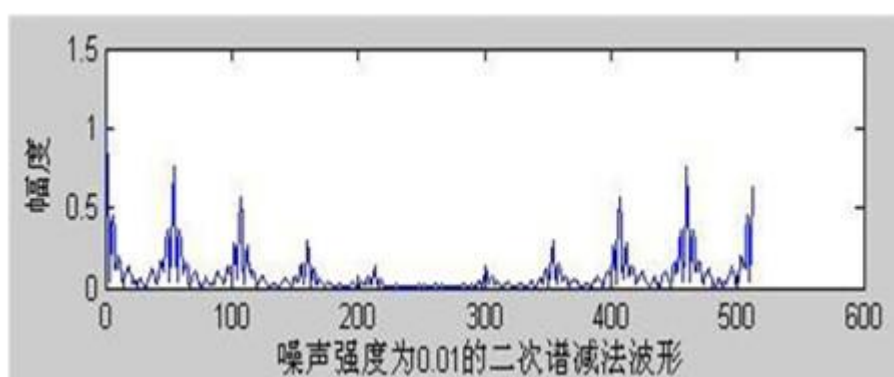
图 21 改进后的基音检测的倒谱法

该方法避免了在倒谱基音检测中求对数,从而减少了所涉及的非线性计算所带来大量的计算,是一种二次谱处理方法,由于浊音线状谱导致了其二次谱具有周期性,因此,在其二次谱频率轴上除零频外的某一特定点将会出现一个最大谱峰。高斯白噪声 W_n 在作二次谱处理后是不具有这些性质的。因此,在理论上通过估计这个二次谱频率不为零的最大谱峰值的横坐标值即可测出浊音的周期。

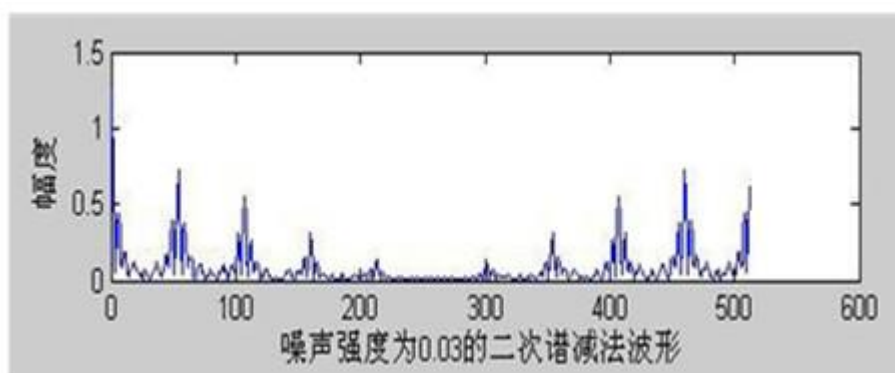
通过上面的分析可知倒谱法进行基音检测主要围绕领域进行,时域的变化对其影响不大,故本次实验就分析二次谱减法的抗噪性能。总共进行四次试验:第一次是用原始信号不加任何处理进行二次谱减分析;第二次是把原始信号加强度分别为 0.01 噪声后,进行分析;第三次是把原始信号加强度分别为 0.03 噪声后,二次谱减分析。第四次是把原始信号加强度分别为 0.1 噪声后,二次谱减分析。实验结果如下:



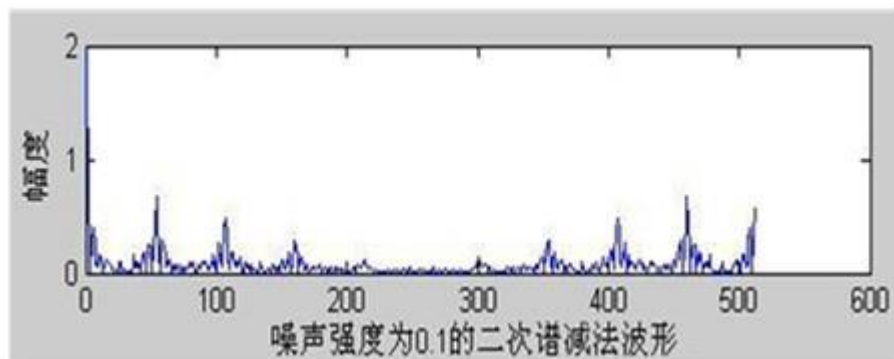
(a) 原信号二次谱减法分析



(b) 加噪(强度为 0.01)后的二次谱减法分析



(d) 加噪(强度为 0.03)后的二次谱减法分析



(e) 加噪(强度为 0.1)后的二次谱减法分析

图 22 二次谱减法基音检测仿真图

此段浊音的基音周期计算为： $T_1 = (0.1 * 58) / 1000 = 5.8\text{ms}$ ，

同理可以估算出不同的噪声环境下估算的基音周期分别为 $T_2 = 6.2\text{ms}$ ，

$T_3 = 7.5\text{ms}$ ， $T_4 = 8.4\text{ms}$ 。

4.1.4 短时自相关函数法与倒谱法比较分析

本文用 PRAAT 软件测出的基音周期为基准，PRAAT 是阿姆斯特丹大学语音研究中用电脑做语音分析的专用软件,是一款相当不错的共享语音分析软件，因为它有着较强的语音分析和数据处理能力，故由其测出的基音周期具有一定的准确性，测得 $T = 5.5\text{ms}$ 。

由实验数据可以得到用自相关法和倒谱法以及改进后的二次谱减法分别测得语音信号周期的准确度，为了便于观察现将计算结果以表格的形式给出，如下所示：

表 1 自相关法与倒谱法的比较

次数	仿真方法		带通滤波	中心削波	噪声强度	标准周期(ms)	实验结果(ms)	准确度(%)
	自相关	倒谱法						
1	是		否	否	无	5.5	6.0	90.1
2	是		是	否	无	5.5	5.8	94.5
3	是		否	是	无	5.5	5.9	92.7
4	是		否	否	0.05	5.5	6.2	87.3
5	是		否	否	0.2	5.5	7.8	76.4
6	是		否	否	0.5	5.5	无法测出	0
7		是	否	否	无	5.5	5.7	96.4
8		是	否	否	0.01	5.5	6.0	90.1
9		是	否	否	0.03	5.5	7.3	85.5
10		是	否	否	0.1	5.5	无法测出	0

表 2 倒谱法与二次谱减法的比较

次数	仿真方法		噪声强度	标准周期(ms)	实验结果(ms)	准确度(%)
	倒谱法	二次谱减法				
1	是		无	5.5	5.7	96.4
2	是		0.01	5.5	6.0	90.1
3	是		0.03	5.5	7.3	85.5
4	是		0.1	5.5	无法测出	0
5		是	无	5.5	5.6	98.2
6		是	0.01	5.5	5.9	92.3
7		是	0.03	5.5	6.5	81.2
8		是	0.1	5.5	7.2	78.2

图 23 仿真计算结果图

由上述计算结果可得出以下结论：

(1) 经低通滤波或中心削波后等预处理的短时自相关法提高了基音检测的准确性;

(2) 倒谱法进行基音检测的准确度比短时自相关法要高,但是抗噪性能不强;

(3) 改进后的二次谱减法比倒谱法的抗噪性能要好,在相同噪声环境下能更加精确地检测出语音信号的基音周期;

(4) 语音信号的时域分析具有简单直观、清晰易懂等优点,但是因为语音信号最重要的感知特性反映在其功率谱中,相位变化只起很小的作用,所以更为有效的分析应围绕频域进行分析。

5 结论与展望

5.1 结论

本文以语音信号的基音检测为研究对象,对短时自相关函数、短时平均幅度差函数、倒谱法、小波变换法这四种基音检测的方法的原理进行了分析。在此基础上,深入研究了短时自相关函数、倒谱法基音检测的算法。通过实验仿真,发现这两种测量方法的不足;最后对这两种算法进行了改进,增加了一些预处理,不仅能够对一段语音信号进行基音检测,还有效地提高了算法在基音检测时的准确性和抗噪性。

本文运用 MATLAB 软件对各种算法进行了仿真,根据大量仿真结果的波形图,可分析出各种算法在准确性和抗噪性等方面特点。主要取得了如下成果:

①分析了各种语音信号基音检测的原理,并用 MATLAB 实验仿真;

②深入分析短时自相关函数法、倒谱法基音检测的算法,实现了对一帧语音信号较准确的基音检测;

② 对经典算法的不足之处,本文提出了改进方法,并取得了较好的效果。

5.2 展望

基音周期的大小与个人的发音习惯及性别、年龄、发音时的力度及情感有很大的关系。传统的基音检测都是以“冷漠的”语音为研究对象的,而在实际中,人在发音时,常常带着惊讶、喜悦、紧张、愤怒、悲伤等复杂的感情。有时环境噪声引起 Lombard

效应(说话人在高噪声环境中为了使自己的声音能被听到,有意或无意地提高嗓音,致使其发音速度及质量发生变化)。在这些感情和力度的驱使下,人的声带开启到最大面积时间及关闭的速度与平静时相比,都要发生变化。同时,共振峰的带宽和位置都要发生畸变,这些变化不可避免地影响到基音周期。

可见不同的情感的影响是不同的,这些影响无疑使上述的基音检测方法的预处理及后处理变得复杂,国外的学者采用实时监控情感的变化,并利用相应的影响来修正基音轨迹,以平滑或动态改变窗的宽度,可以有效降低上述影响。近年来,基于线性预测分析频谱的 Mel 倒谱系数在处理带情感的语音识别中取得很大的进展,能否将此种方法引入到相应状态下的基音检测,仍值得广大学者去研究。

参 考 文 献

- [1]冯康,时慧琨,语音信号基音检测的现状与展望 [J],微机发展,2004.3.
- [2]韩纪庆,张磊,郑铁然.语音信号处理 [M].清华大学出版社,2004.9.
- [3]杨行峻,迟惠生,等.语音信号数字处理 [M].北京:电子工业出版社,1995, 6-8, 29-32.
- [4]鲍长春.低比特率数字语音编码基础 [M].北京:北京工业大学出版社,2001.3.
- [5]赵建伟,楼红伟,徐大为,刘重庆.噪声环境下的基音检测方法 [J].红外与激光工程 2002,21(1): 5-8.
- [6]刘丽娟,冯宏伟,基音周期的一种检测方法 [J],科学技术与工程,2006.9.
- [7]冯康,时慧琨,一种参考短时平均能量和平均幅度差函数的基音检测方法 [J],淮南师范学院学报,2003.5: 9~11.
- [8]王都生,铁满霞,樊昌信.一种实时基音检测算法 [J].电子学报,2000, 28(10): 9~11
- [9]郑燕琳.汉语连续语音三字词基音提取及声调识别 [J].吉林大学硕士学位论文.长春:吉林大学,2004
- [10]Rabiner L R. On the use of autocorrelation analysis for pitch detection [J]. IEEE Trans. Asssp, 1977,25:24-33.
- [11]Ross M J, haffer HL, Cohen A, Freudberg R, et al. Average magnitude difference function pitch extractor [J]. IEEE Trans. Assp, 1974, 22:353-362.
- [12]Ingrid Daubechies Orthonormal bases of compactly supported wavelets[J]. Comm PureAPP1Math, 1988, 11(41):909-996.
- [13]李建平,杨万余译.Ingrid Daubechies 小波十讲[M].北京:国防工业出版社,2004.
- [14]杨树功.小波分析在语音信号基音检测中的应用研究 [J].西北工业大学硕士学位论文.西安西北工业大学,2005.

- [15]李弼程,罗建书. 小波分析及其应用 [M].北京:电子工业出版社. 2003.
- [16]RAB INER LR, CHENGMJ, ROSENBERG AE, et al. A comparative performance study of several pitch detection algorithms [J]. IEEE Transactions on ASSP, 1976, 24 (5) :399 -418.
- [17]马祥杰, 孟相加, 张百生.一种自相关基音检测算法[J].通信技术, 2003.5.
- [18]王和平, 康景利, 基于形态滤波和小波变换的基音检测新方法[J].语音技术, 2005.5.
- [19]华国刚, 戴陪倩, 张钦, 李辉.基于小波变换和平均周期鲁棒性基音周期检测方法[J].数据采集与处理, 2002.3.
- [20]L. R. Rabiner and R. W. Schafer.朱雪龙等译[J]. 语音信号数字处理.北京:科学出版社, 1983.
- [21]宋兵.语音信号的基音检测及其应用[J].上海大学硕士学位论文.上海:上海大学, 2005.
- [22]付青青, 吴爱平.基于 Matlab 语音信号自相关基音检测[J]. 长江大学学报.2006,3(4) :99-101.
- [23] 吴谨, 刘振兴. 基于小波变换的语音基频检测.武汉科技大学学报 [J] . 武汉: 武汉科技大学, 2000, 23(2) :8-11.
- [24]许钢, 黄冰. 基于小波变换和归一化自相关的基音检测算法 [J] . 桂林电子工业学院学报. 2003. 23:1-5.
- [25]赵军. 基于独立分量分析和基音提取技术的盲源分离[J]. 西安理工大学硕士学位论文. 西安:西安理工大学, 2005.
- [26]董靖. 基于语音增强及小波变换基音检测的 LPC-MBE 低速语音编码算法[J].吉林大学硕士学位论文. 长春:吉林大学,2004.
- [27]Parsons T W.Voice and Speech Processing[J]. New York:McGraw-Hill BookCompany, 1986.
- [28]G.Fant. Acoustic theory of speech production[J]. The Hague (TheNether-lands) :Mouton and Co. 1960.
- [29]Bing Xiang and Toby Berger. Efficien. text-independent speaker verification with structural gaussian mixture models and neural network [J] .IEEE Transaction on Speech and Audio Processing. 2003, 11(5) :447-456.
- [30]胡航. 语音信号处理[M]. 哈尔滨: 哈尔滨工业大学出版社, 2000: 116-126
- [31]L. R. Rabiner. on the use of autocorrelation analysis for pitch detection[J]. IEEE Trans. on Acoustics Speech and Signal Proc, 1 977, 26(1): 24-33.
- [32]Ross M J, Shaier H L, Cohen A, et al. Average magnitude difrerence function pitch extractor[J]. IEEE Trans on Acoustics Speech Signal Processing, 1 974, 22(5): 353-362.
- [33]Markel J D. The SIFT algorithm for fundamental frequency estimation[J]. IEEE Trans on Audio and Elect, 1 972, 20(5): 367—377.
- [34]A. M. Noll. Cepstrum pitch determination[J]. J. Acoust. Soc. Amer. 1967, 41(2): 293—309.
- [35]]陈小利. 含噪语音基音周期检测算法研究[D]. 解放军信息工程大学, 2007.

致谢

本论文是在导师钱何朝霞讲师精心指导和悉心关怀下完成的。从论文的选题、定题，论文研究内容的探讨，到论文大纲、确定稿的形成，期间每一个阶段都渗透着老师的心血。导师严谨的治学态度、踏实的工作作风、锐意进取的开拓精神、追求科学前沿的求实学风、乐观自信的风度、高尚朴实的人格，这也潜移默化地影响着我，也将指引我未来工作和生活前进的道路。在此，怀着深深的敬意，谨向导师致以最真挚的谢意！

特别感谢我的父母，感谢他们和风细雨般的教诲，感谢他们百般呵护、无微不至的关怀。多年来的含辛茹苦，他们为了我的成长倾注了太多的心血和无私的关爱。不平坦的求学路上，父母用他们温暖宽厚的双手搀扶着我一路走来，不管遇到多大的雨雪风霜，只要有父母在身边，一切都会迎刃而解，他们永远都是我坚强的后盾和避风的港湾。亲情似海，恩重如山！

感谢为审阅本文而付出辛勤劳动的所有老师们。

最后，衷心感谢所有关心和帮助我的人，祝福他们健康、幸福、平安！

附录：

程序清单一（短时自相关函数法基音检测）：

```
Y=WAVREAD('pitch.wav',);
subplot(211);
plot(x);
grid;
xlabel('原始信号');ylabel('幅度')
x=Y(500:799);%取帧
x=x.*boxcar(300);
R=zeros(1,300); % 自相关分析
for k=1:300
    for n=1:300-k
        R(k)=R(k)+x(n)*x(n+k);
    end
end
j=1:300;
subplot(212);
plot(j,R);
grid;
xlabel('浊音信号的自相关函数');ylabel('幅度')
```

```
Y=WAVREAD('pitch.wav');
x=Y(500:799);%取帧
x=x.*boxcar(300);
R=zeros(1,300); % 自相关分析
for k=1:300
    for n=1:300-k
        R(k)=R(k)+x(n)*x(n+k);
    end
end
j=1:300;
subplot(211);
plot(j,R);
grid;
xlabel('原始信号的自相关函数');ylabel('幅度')
x=x+0.05*randn(size(x));%加噪
x=Y(500:799)
x=x.*boxcar(300);
R=zeros(1,300); % 自相关分析
for k=1:300
    for n=1:300-k
```



```

        R(k)=R(k)+x(n)*x(n+k);
    end
end
j=1:300;
subplot(212);
plot(j,R);
grid;
xlabel('加噪信号的自相关函数');ylabel('幅度')

```

程序清单二（短时平均幅度差函数法基音检测）：

```

Y=WAVREAD('pitch.wav');%读取语音信号
x=Y(500:799);%取帧
x=x+0.05*randn(size(x));
subplot(211);
plot(x);
grid;
subplot(211)
xlabel('加噪 S(n)');ylabel('幅度')
x=x.*boxcar(300);
R=zeros(1,300); %ADMF 分析
for k=1:300
    for n=1: 300-k
        R(k)=R(k)+abs(x(n+k)-x(n));
    end
end
subplot(212)
j=1: 300;
plot(j,R);
grid;
xlabel('加噪后信号的 AMDF 函数');ylabel('幅度')

```

程序清单三（倒谱法基音检测）：

```

[signal,Fs]=wavread('pitch.wav');%读取语音号
Y=signal(500:1011);%选择其中 512 点
Y=Y.*hamming(512);%加汉明窗
a=fft(Y,512);%%FFT 变换
b=a.*conj(a);%
b=sqrt(b);%求模
c=log(b);%取对数
subplot(211)
plot(c);

```

```
xlabel('对数幅度谱');ylabel('幅度')
d=ifft(c,512);%逆变换
subplot(212);
plot(abs(d));
xlabel('倒谱波形');ylabel('幅度')
```

程序清单四（经中心消波处理后自相关函数法基音检测）：

```
y=WAVREAD('pitch.wav');
x=y(500:799);
x=x.*boxcar(300);
R=zeros(1,300);
for k=1:300
    for n=1:300-k
        R(k)=R(k)+x(n)*x(n+k);
    end
end
j=1:300;
subplot(211);
plot(j,R);
grid;
xlabel('原始信号的自相关函数');ylabel('幅度')
for i=500:799;
    if ( y(i)>0.5),
        x(i)=y(i)-0.5;
    elseif (y(i)<-0.5),
        y(i)=y(i)+0.5
    elseif (abs(y(i))<=0.5),
        y(i)=0;
    end
end
x=y(500:799);
subplot(211);
plot(x);
grid;
xlabel('经中心削波处理的信号');ylabel('幅度')
x=x.*boxcar(300);
R=zeros(1,300);
for k=1:300
    for n=1:300-k
        R(k)=R(k)+x(n)*x(n+k);
    end
end
end
```

```
j=1:300;
subplot(212);
plot(j,R);
grid;
xlabel('经过中心消波处理后的信号自相关函数');ylabel('幅度')
```

程序清单五（小波变换）：

```
fs=wavread('pitch.wav');
s=fs(500:799);
subplot(311);
plot(s);
grid;
xlabel('原始信号');
% s=s.*hamming(300);
[c,l]=wavedec(s,16,'db20');
a6=wrcoef('a',c,l,'db20',6);
subplot(312);
plot(a6);
grid;
xlabel('小波变换第 5 层低频信号')
```

```
d6=wrcoef('a',c,l,'db20',6);
subplot(313);
plot(d6);
grid;
a5=wrcoef('d',c,l,'db20',5);
subplot(313);
plot(a5);
grid;
xlabel('小波变换第 5 层高频信号')
```

程序清单六（低通滤波后自相关）：

```
b1=fir1(48, 0.09);
Y=WAVREAD('pitch.wav')
x=Y(500:799);
t=1;1/10000;1/100;
Y=filter(b1,1,x);
subplot(211);
plot(Y);
grid;
xlabel('低通滤波后的信号');ylabel('幅度')
```

```
b1=fir1(48, 0.09);
Y=WAVREAD('pitch.wav')
t=1;1/10000;1/100;
Y=filter(b1,1,Y);
x=Y(500:799);
x=x.*boxcar(300);
R=zeros(1,300);
for k=1:300
    for n=1:300-k
        R(k)=R(k)+x(n)*x(n+k);
    end
end
j=1:300;
subplot(212);
plot(j,R);
grid;
xlabel('低通滤波后的自相关函数');ylabel('幅度')
```

程序清单七（二次谱减法）：

```
Y=WAVREAD('pitch.wav');%读取语音号
Y=Y(500:1011)
Y=Y.*hamming(512);%加汉明窗
subplot(311);
plot(Y);
a=fft(Y,512);%%FFT 变换
b=a.*conj(a);%
b=sqrt(b);%求模
c=ifft(b,512);%逆变换
b=c.*conj(c);%
b=sqrt(b);%求模
subplot(313);
plot(abs(b));
xlabel('二次谱减法波形');ylabel('幅度')
```

程序清单八（窗函数）：

```
n=64;
w1=boxcar(n);
subplot(211);
plot(w1);
xlabel('采样点数');
ylabel('幅度');
[w1,f]=freqz(w1/sum(w1),1,512,2);
```

```
subplot(212);  
plot(f,20*log10(abs(w1))),grid;  
xlabel('采样点数');  
ylabel('幅度/dB');
```