目录

[第一章 绪论 1](#_Toc357026451)

[1.1引言 2](#_Toc357026452)

[1.2语音的特点 2](#_Toc357026453)

[1.3人耳的感知特性 2](#_Toc357026454)

[1.4噪声的特点 3](#_Toc357026455)

[1.5常见语音增强消噪算法简介 4](#_Toc357026456)

[1.6本文的主要工作 4](#_Toc357026457)

[第二章 语音信号的预处理 5](#_Toc357026458)

[2.1引言 5](#_Toc357026459)

[2.2语音信号的采集 5](#_Toc357026460)

[2.3语音信号的预加重处理 5](#_Toc357026461)

[2.4语音信号的分帧处理 7](#_Toc357026462)

[第三章 声音信号的短时处理 9](#_Toc357026463)

[3.1 引言 9](#_Toc357026464)

[3.2 短时平均幅度函数 9](#_Toc357026465)

[3.3 短时平均能量 9](#_Toc357026466)

[3.4 短时平均过零率 10](#_Toc357026467)

[3.5 短时自相关分析 10](#_Toc357026468)

[第四章 噪声估计 12](#_Toc357026469)

# 

# 第一章 绪论

## 1.1引言

语音是人类相互之间使用最多,也是重要的信息交互手段。一个实用的语音系统，如语音编码系统，语音识别系统等，在设计制造完成后处于现实工作环境中时，将遇到各种各样的环境噪声，比如公众场合中对手机讲话，手机中麦克风会将各种背景噪声与该用户的讲话一起，同时录入作为后续语音编码或者识别系统的输入。有时候此类噪声将使得在实验室环境下性能良好的语音系统性能恶化，甚至工作完全失效，因此设计语音系统时，必须考虑其工作环境的影响。现实中噪声往往不可避免，所以需要我们采取各种措施，提高语音系统在噪声环境下的工作性能。

在这种情况下,语音增强的目标是从带噪声的语音信号中提取尽可能纯净的原始语音,抑制背景环境噪声,提高语音的质量并提高听者的舒适度,使听者不感觉疲劳。另外语音识别正进入实用阶段,语音识别的应用也不断扩大,而语音识别中抗噪声干扰是提高语音识别率的一个重要环节。这时语音增强的目的关键是加强语音识别特征,使语音易于识别。语音增强是一门涉及面很广的交叉学科,它不但与语音信号数字处理理论有关,而且涉及到模式识别!数理统计!神经生理学和语音学学科。此外,语音增强所要面临的噪声来源也可能众多,常见的如街道、机场、工厂车间、人声嘈杂的公共场合等。因此,人们在研究语音增强的方法时一般是从语音特点、人耳的感知特性和噪声特点等几方面入手。

## 1.2语音的特点

语音信号是一种非平稳的随机信号,同时也可以看作是一种短时的平稳的随机信号。这是因为人类发声过程的变化速度是有一定的限度,一般在短时间内(10-30ms)人的声带与声道的形状有相对的稳定性,在这段时间里认为语音是物理特性与频谱特性近似不变的。语音的短时特性是语音信号分析和处理的基础,利用这一特性就可以应用平稳随机过程来分析与处理语音信号。

任何语音都包含元音与辅音两种音素,辅音根据声带是否振动又可分为清辅音与浊辅音两种。浊辅音在时域上呈现出明显的周期性和较强的振幅,在频域上有共振峰结构,而且能量大部分集中在较低频段内。清辅音则完全不同,它没有明显的时域和频域特征,类似于白噪声并振幅较弱。语音增强中可以利用浊音具有明显的周期性来区别和抑制非语音噪声,而清辅音的特性则难与宽带噪声区分。根据中心极限定理,傅里叶展开系数被认为是独立的高斯随机变量,均值为零,而方差是时变的。这种高斯模型应用于有限帧长时只是一种近似的描述,在宽带噪声污染的带噪语音的语音增强中,这种假设可做为分析的前提。

## 1.3人耳的感知特性

语音感知问题涉及到生理学、心理学声学和语音学等诸多领域,这是一个复杂的问题,有待进一步研究。尽管如此目前已有一些有用的结论可以应用于语音增强:

(l)人耳对语音的感知是通过语音信号中各频谱分量幅度获取的,对各分量的相位则不敏感。

(2)人耳对频谱分量强度的感受是频率与能量谱的二元函数,响度与频谱幅度的对数成正比。

(3)人耳对频率高低的感受近似与该频率的对数值成正比。

(4)人耳有掩蔽效应即强信号对弱信号有掩盖的抑制作用。掩蔽的程度是声音强度与频率的二元函数。对频率临近分量的掩蔽要比频差大的分量有效得多。

(5)短时谱中的共振峰对语音的感知十分重要,特别是第二共振峰比第一共振峰更为重要,因此对语音信号进行一定程度的高通滤波不会对可懂度造成影响。

(6)人耳在两人以上的讲话环境中有能力分辨出需要聆听的声音。

## 1.4噪声的特点

噪声可以是加性的,也可以是非加性的。对于非加性噪声,有些可以通过变换而转变为加性噪声。这里仅讨论加性噪声。加性噪声大致有:周期噪声!脉冲噪声!宽带噪声和同声道其他语音千扰。

(l)周期性噪声。它的特点是在频域上具有许多离散的线谱。周期性噪声主要来源于周期性转动的机械和电气干扰。

(2)脉冲噪声。脉冲噪声表现为时一域波形中出现的窄脉冲。它来源于爆炸,撞击和放电等。

(3)宽带噪声。它的来源有很多种,如热噪声,气流(如风,呼吸)噪声及各种随机噪声源。由于宽带噪声与语音信号在时域和频域上都完全重叠,因而消除它最为困难。这种噪声只有在语音间隙才单独存在,对于平稳的宽带噪声通常可以认为是高斯白噪声。对不具有白色频谱的噪声,可以先进行白化处理转化为白噪声。

(4)同声道语音干扰。人耳可以在两人以上讲话环境中分辨出所需要的声音,这种分辨能力是人体内部语音理解机理的一种感知能力。这种能力来源于人的双耳输入效应,但当多个语音叠合在一起,在单信道中传输时,双耳信号因合并而消失。

## 1.5常见语音增强消噪算法简介

由于噪声来源众多,随应用场合而异,它们的特性也各不相同,难以找到一种通用的语音增强算法适用于各种噪声环境。而且语音增强不但与语音信号数字处理理论有关,还涉及到人的听觉感知和语音学,所以必须针对不同的噪声,采取不同的语音增强对策。几十年来人们在语音增强方面做了很多不懈的探索,总结出适应不同情况的各种增强方法。

1. 周期噪声：陷波滤波器（梳状滤波器、FFT滤波器）。
2. 脉冲噪声：根据其持续时间段的特点抛弃某段语音；或者设置削波器，减弱其影响。
3. 宽带噪声：谱减法参数谱估计法、Wiener滤波器法、盲信号分离法。
4. 同声道语音干扰：基音频率增强法、麦克风阵列增强法。

## 1.6本文的主要工作

本文的重点是研究现有的语音增强方法，实现一种实用的增强算法。本文的主要工作包括：

1. 参阅国内外相关文献资料，介绍语音增强消噪技术的背景与常见方法。
2. 研究常用的噪声估计方法。
3. 用陷波滤波器，消波器，谱减法及基音频率增强法实现一个完整的语音增强及消噪系统。

# 第二章 语音信号的预处理

## 2.1引言

实际的语音信号代表了空气的密度随时间的变化，基本上是一个连续的函数，但是若要将此信号储存在计算机里，就必须先将此信号数字化。另外，为了增加语音的高频分辨率及提取音频特征，我们要对信号进行预处理。

## 2.2语音信号的采集

一般而言，当我们将声音储存到电脑时，有下列几个参数需要考虑：

1. 采样频率（Sample Rate）：每秒种所取得的声音资料点数，以赫兹（Hz）为单位，采样频率越高声音品质越好，但是数据量越大。常用采样频率如下：

1. 8KHz：电话及一般玩具内语音芯片的音质

2. 16KHz：语音识别系统的音质

3. 44.1KHz：CD音质

实际采集音频时使用的频率依照奈奎斯特采样定理来确定。

（2） 采样解析度（Bit Resolution）：存储采样后的数字信号时使用的位元数，常用的数值如下：

1. 8-bit：可表示的数值范围为0~255 或-128~127

2. 16-bit：可表示的数值范围为-32768~32767

换句话说，每个取样点的数值都是整数，以方便储存。但是在MATLAB 中，通常把音讯的值正规化到[-1, 1] 范围内的浮点数，因此若要转回原先的整数值，就必须再乘上2^nbits/2，其中nbits 是采样解析度。

1. 声道：一般只分单声道（Mono）或立体声（Stereo），立体音即是双声道。

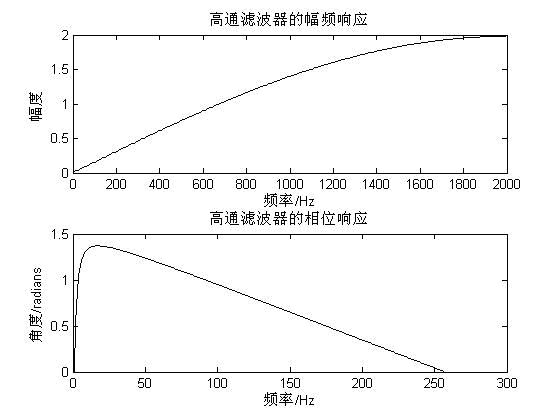
## 2.3语音信号的预加重处理

语音信号的功率谱随频率的增加而减小，其能量大部分集中在低频部分，这就可能造成高频部分的信噪比降低到不能容许的程度。为了抵消这种现象，需要在音频处理之前对信号进行预加重处理。

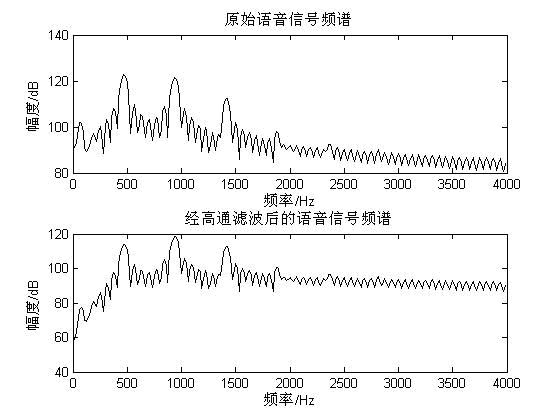
可以通过一阶FIR高通数字滤波器来实现预加重：

设n时刻的语音采样值为x（n），经过预加重处理后的结果为：

这里取0.98，下图为高通滤波器的幅相频特性曲线：



预加重前后的一段音频信号：



## 2.4语音信号的分帧处理

由于发音器官的惯性运动，可以认为在一小段时间里（一般为10ms~30ms）语音信号近似不变，即语音信号具有短时平稳性。这样，可以把语音信号分为一些短段（称为分析帧）来进行处理。

音频信号分帧的基本原则如下：

（1）将音频切成一个个音框，音框长度大约是20~30 ms。音框若太大，就无法抓出音讯随时间变化的特性；反之，音框若太小，就无法抓出音频的特性。一般而言，音框必须能够包含数个音讯的基本周期。（另，音框长度通常是2 的整数次方，若不是，则在进行傅里叶变换时，需补零至2 的整数次方，以便使用快速傅里叶变换）

（2）若是希望相邻音框之间的变化不是太大，可以允许音框之间有重叠，重叠部分可以是音框长度的1/2 到2/3 不等。（重叠部分越多，对应的计算量也就越大。）

信号分帧时的常用名词为：

音框点数（Frame Size）：每一个音框所含有的点数。

音框重叠量（Frame Overlap）：音框之间重叠的点数。

音框跳距（Frame Step or Hop Size）：此音框起点和下一个音框起点的距离点数，等于音框点数减去音框重叠。

音框率（Frame Rate）：每秒出现的音框数目，等于取样频率除以音框跳距。

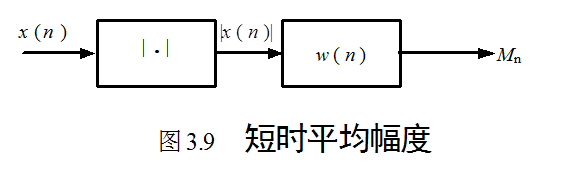
# 第三章 声音信号的短时处理

## 3.1 引言

由于发音器官的惯性运动，可以认为在一小段时间里（一般为10ms~30ms）语音信号近似不变，即语音信号具有短时平稳性。因此，我们可以对信号进行短时处理。

## 3.2 短时平均幅度函数

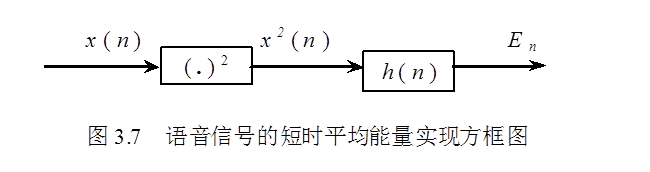
短时平均幅度为一帧内各点值的绝对值之和，定义为：



我们可以使用短时平均幅度来描述音频信号某帧的能量。此方法计算简单，适用于计算能力不强的平台。

## 3.3 短时平均能量

为了使能量描述更加符合人耳的主观感受，我们定义短时平均能量定义为：



短时平均能量实现方框图

这种方法得到的值以分贝为单位，比较符合人耳的主观感受。此法计算较为复杂，适用于高性能计算机。

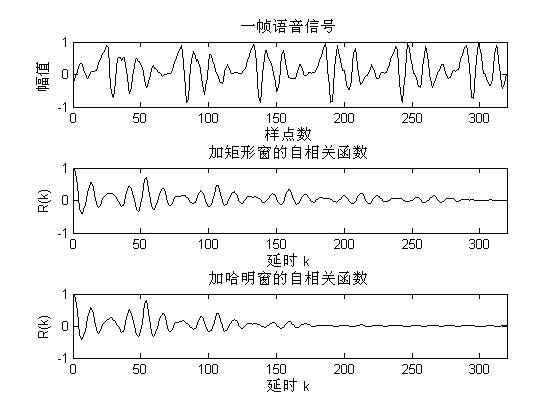
## 3.4 短时平均过零率

在离散时间语音信号情况下，如果相邻的采样具有不同的代数符号就称为发生了过零。单位时间内过零的次数就称为过零率。过零率定义为：

一般而言，噪音及清音的过零率均大于浊音。短时平均过零率通常用在语音端点检测，特別是用在估测清音的起始位置及结束位置。

## 3.5 短时自相关分析

时域离散确定信号的自相关函数定义为：



浊音的自相关函数

自相关函数有性质：k=0时自相关函数取得最大值，所以对于周期为P的信号，在取样0，P，2P，…处，其自相关函数也是最大值，因此，可以根据自相关函数的最大值位置来估计周期信号的周期值。

在语音信号处理中，我们可以用自相关函数来判断一段音频的音高。

# 第四章 噪声估计