

语音与音频工程

南京大学 电子科学与工程学院

撰稿人：方元

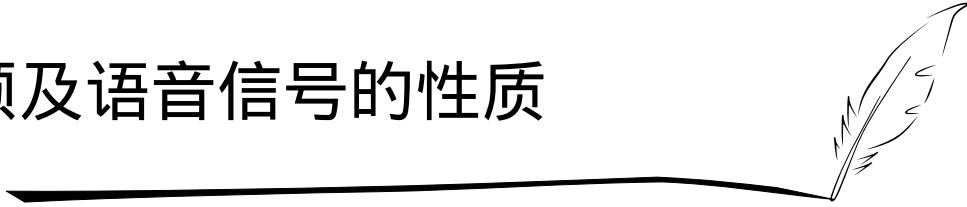
2022 年 8 月

目录

1. 音频及语音信号的性质	1
1.1. 声波的物理特性	1
1.2. 听觉系统	2
1.2.1. 振动感知	2
1.2.2. 掩蔽效应	5
1.2.3. 鸡尾酒会效应	5
1.2.4. 响度	5
1.3. 语音的基本特征	6
1.3.1. 语音的传输过程	6
1.4. 常用音频仪器	7
1.4.1. 音频设备测量	7
1.4.2. 声音采集和还原系统	8
2. 语音参数分析	9
2.1. 语音产生的生理模型	9
2.1.1. 音位	9
2.1.2. 发音器官	10
2.2. 语音频谱特征	10
2.2.1. 音调	11
2.2.2. 共振峰	11
2.3. 短时分析	12
2.3.1. 短时语音参数	13
2.3.2. STFT	14
2.4. 线性预测编码	15
2.5. 同态信号处理	16
3. 数字滤波系统	19
3.1. 有限冲激响应滤波	19
3.1.1. 理想滤波器模型	19
3.1.2. 冲激响应计算	19

3.1.3. 滤波算法	24
3.2. 无限冲激响应滤波	25
3.2.1. 零极点配置法	25
3.2.2. 由模拟滤波器变换	26
3.2.3. 滤波器类型变换	31
3.2.4. 峰值滤波/陷波	34
3.2.5. IIR 滤波器结构选型	35
3.3. 梳状滤波器	37
4. 自适应滤波器	38
4.1. 维纳滤波	38
4.1.1. 问题的提出	38
4.1.2. Wiener-Hopf 方程	40
4.2. 自适应滤波	41
4.2.1. 最速下降法	41
4.2.2. 最小均方自适应滤波算法	41
4.2.3. 自适应滤波的应用	43
5. 音频工程应用	46
5.1. 工程计算中的定点小数处理	46
5.2. 单频信号谱分析	47
5.3. 双音多频检测	48
5.4. 回声与混响	49
5.5. 变声	51
5.6. 压限	52
5.7. 阵列信号处理	53
5.7.1. 多路输出系统	53
5.7.2. 阵列扬声器	53
5.7.3. 阵列传声器	54

音频及语音信号的性质



1.1 声波的物理特性

声波是一种机械波, 它需要传播介质。我们通常所说的声音是指在空气中传播的声波。

在空气中, 声音大小是通过压强表示的。在绝对无声的环境中, 空气的压强为 1 个标准大气压力 1 atm (等于 101.325 kPa)。当有声波存在时, 空气分子随声波振动, 导致空气压强在原有大气压的基础上产生微小的变化。这个微变的压强就是声压。

流体中支配声波运动的三个基本方程: 运动方程, 连续性方程和状态方程都是非线性的。以静态的理想流体为例, 声压 p 与流体质点速度 \mathbf{v} 服从如下方程

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) p = -\kappa_s \nabla \cdot \mathbf{v} \quad (1.1)$$

$$\rho \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v} = -\nabla p \quad (1.2)$$

其中 ρ 为瞬时流体质量密度, κ_s 为瞬时媒质体弹性系数。

根据声学理论建立起的声波波动方程是一个非线性微分方程。当声压 p 和速度 v 变化量均是微小量时, ρ 和 κ_s 可分别用静态常数值 ρ_0 和 κ_{s0} 代替, 并忽略方程中的非线性项, 从而得到线性化的方程

$$\frac{\partial p}{\partial t} = -\kappa_{s0} \nabla \cdot \mathbf{v} \quad (1.3)$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla p \quad (1.4)$$

由此可得标准的线性波动方程

$$\nabla^2 p - \frac{\rho_0}{\kappa_{s0}} \frac{\partial^2 p}{\partial t^2} = 0 \quad (1.5)$$

声速 $c = \sqrt{\kappa_{s0}/\rho_0}$ 。介质的密度越低, 声速越大; 介质的刚性越大, 声速也越大。

声线性方程的解和电磁波方程的解的特性非常相似, 它也具有叠加特性、干涉特性。大多数音频处理都是基于线性理论而开展的。然而, 声波的非线性特性也会导致非常有趣的现象。非线性问题主要出现在超声领域, 许多声学研究人员正在从事这方面的研究。

表 1.1: 各种声音对应的声压级

空气中的声源	声压级 SPL (dB)
激波 (超音速)	>194
非失真声音的理论极限	~194
闪光弹	170 – 180
火箭发射	~165
1 米外来福枪射击声	171
1 米外喷气机引擎	150
痛感阈值	130 – 140
人声极限	135
听力损伤	120
30–100 米外的喷气机引擎	110 – 140
1 米外的手提钻	100
10 米外交通繁忙的公路	80 – 90
长期暴露的听力损伤下限	85
10 米外汽车行驶声	60 – 80
听力保护上限 (美国环保局)	70
洗衣机、洗碗机	50 – 53
1 米距离的正常对话	40 – 60
非常安静的房间	20 – 30
树叶声、安静的呼吸声	10
人的听觉下限	0
微软的消声室	-20

人耳的可听域相当宽, 以声压表示的可听范围可以超过好几个数量级。通常又将声压转换成分贝 (dB) 表示, 称为声压级。人耳最灵敏的频率 (2kHz~5kHz) 处可听到的最小声压是 $20\mu\text{Pa}$, 于是就把 $20\mu\text{Pa}$ 定义为 0dB。

对于点声源, 声压与距离成反比, 即:

$$p \propto \frac{1}{r}$$

1.2 听觉系统

1.2.1 振动感知

图1.1是人耳的内部结构示意图。声音处理部分分三层结构: 一、外耳, 由耳廓收集声音并通过外部传导通向中耳的管道; 二、从鼓室开始的中耳膜或鼓膜, 它负责将空气中的声波转换成机械压力波; 三、内耳, 由耳蜗和一组与听觉的神经连接组成, 它将神经信号传导到大脑。图1.2是

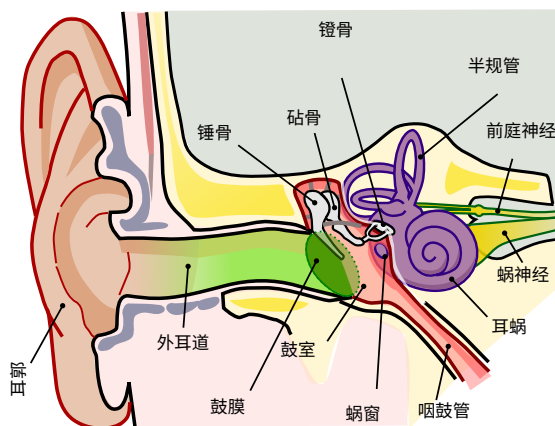


图 1.1: 人耳的生理结构

等效的信号传递模型。

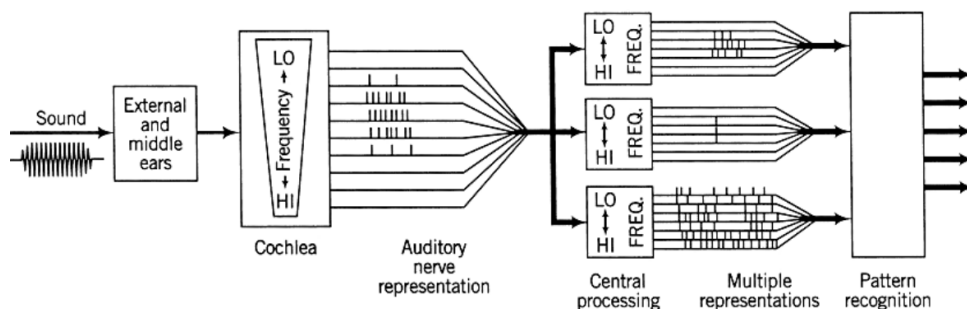


图 1.2: 听觉系统的信号传递模型

声波从外耳传到耳鼓和骨骼结构转换声音的内耳波到机械振动，最终转移到耳蜗内的基底膜 (basilar membrane)。基底膜的不同位置对振动频率的响应差别，表现出一种选频特性，可以等效为由一组带通滤波器构成的频率分析仪。其频率响应随着频率的增加而变得越来越宽。

描述基底膜这种滤波器组的一个重要概念是临界带宽 (critical bandwidth)。由于基底膜上的点彼此牵连，不能独立振动，因此带通滤波器并不理想，相邻带之间有显著的重叠。即便如此，带通滤波器分析的模型也有比较好的应用。实验表明，500Hz 以下，有效带宽大约是 100Hz；高于 500Hz，相对带宽约是中心频率的 20%。一个经验公式是：

$$\Delta f_c = 25 + 75 \left[1 + 1.4(f_c/1000)^2 \right]^{0.69}$$

在 0 到 20kHz 范围内，可划分出约 25 个临界带。

巴克刻度 (又称 Bark 尺度，英语：Bark scale) 是由德国声学家 Zwicker¹ 于 1961 年提出的一种心理声学的尺度，是一种频率的标度。它以第一次对响度作出主观测量的德国物理学家 Barkhausen² 命名。在这个标度上，相等的距离对应于感知上相等的距离。它将人的可听域划

¹ Karl Eberhard Zwicker (1924.1.15 – 1990.11.22) 德国声学专家，慕尼黑工业大学教授。

² Heinrich Georg Barkhausen (1881.12.2 – 1956.2.20)，德国物理学家。1919 年发现了铁磁体的 Barkhausen 效应，1921 年研究出电子振荡器的 Barkhausen 判据。

分为 24 个通带, 它们与听觉的临界频带相对应。在大约 500 Hz 以上这个尺度更相似于一个对数频率轴; 在 500 Hz 以下则更接近线性。

表 1.2: Bark 刻度的中心频率和带宽

Bark	中心频率 (Hz)	截止频率 (Hz)	带宽 (Hz)
1	60	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150
9	1000	1080	160
10	1170	1270	190
11	1370	1480	210
12	1600	1720	240
13	1850	2000	280
14	2150	2320	320
15	2500	2700	380
16	2900	3150	450
17	3400	3700	550
18	4000	4400	700
19	4800	5300	900
20	5800	6400	1100
21	7000	7700	1300
22	8500	9500	1800
23	10500	12000	2500
24	13500	15500	3500

梅尔刻度 (又称 Mel 尺度, 英语: Mel scale) 是一种非线性刻度单位, 表示人耳对等距音高 (pitch) 变化的感受。Mel 的名字来源于单词 melody, 表示这个刻度是基于音高比较而创造的。

梅尔刻度与线性的频率刻度 Hz 之间可以进行近似的数学换算。一个常用的将 f Hz 转换为 m mel 的公式是:

$$m = 2595 \lg \left(1 + \frac{f}{700} \right) \quad (1.6)$$

其参考点定义是将 1000Hz、且高于人耳听阈值 40dB 以上的声音信号, 定为 1000 mel。在频率 500Hz 以上时, 人耳感觉到等量的音高变化, 所需要的频率变化随频率增加而愈来愈大。这样的

结果是, 在赫兹刻度 500Hz 往上的四个八度 (一个八度即为两倍的频率), 只对应梅尔刻度上的两个八度。³

1.2.2 掩蔽效应

临界带引发的一个重要后果是听觉掩蔽效应 (图1.3)。在同一个临界带中, 响亮的音调导致基底膜上某一点的强烈振动, 可以淹没附近频率发生的振动, 从而将弱声淹没在我们的听觉系统之外。纯音可以掩盖其他纯音, 推而广之, 噪音也可以掩盖纯音。

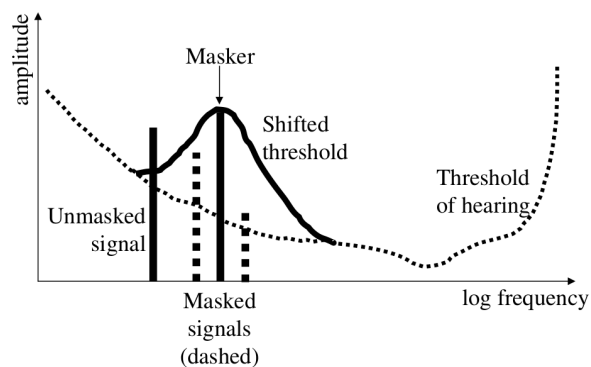


图 1.3: 掩蔽效应示意

除了频域掩蔽特性以外, 时域也存在掩蔽特性。当噪音在欲听到的声音之前, 称为超前掩蔽; 噪音在该声音之后, 称为滞后掩蔽。一般来说, 超前掩蔽很短, 只有 5 到 20ms, 而滞后掩蔽可以持续 50 到 200ms。

1.2.3 鸡尾酒会效应

鸡尾酒会效应 (cocktail party effect) 是指人的一种听力选择能力。最早由英国科学家 Colin Cherry 于 1953 年研究并命名了这一现象。他注意到, 即使在鸡尾酒会这样嘈杂的场合, 对话者可以忽略背景噪声, 仍然能够很好地听清对方的语言。它除了利用双耳的定向能力, 更多的是利用我们听觉系统对声音的分析能力。

1.2.4 响度

响度是感知语音和其他声音的一个重要参数。它以 dB 为单位的纯音声压级为标准参考值, 标度为相同音调的感知响度, 见图 1.4。响度的单位是 phons (昉)。响度曲线表明响度的感知与频率有关。图中实线是通过比较测量的等响度级轮廓, 以 1000Hz 频率的纯音调发出各种频率的声音和已知声压级。例如, 频率为 100Hz 的点在标记为 50 (phons) 的曲线上是调整 100Hz 音调, 直到听起来像 1000Hz 音调一样响亮 50dB 声压级。测量表明, 100Hz 音调必须具有大约 60dB 的声压级才能被感知到与 1000Hz 音调的响度相等压力水平 50 dB。即, 50 dB/1000 Hz 和 60dB/100Hz 音调的响度级别为 50 昉。

³ 基于听觉频率近似对数关系的心理声学研究, 不同学者给出过不同的实验公式。

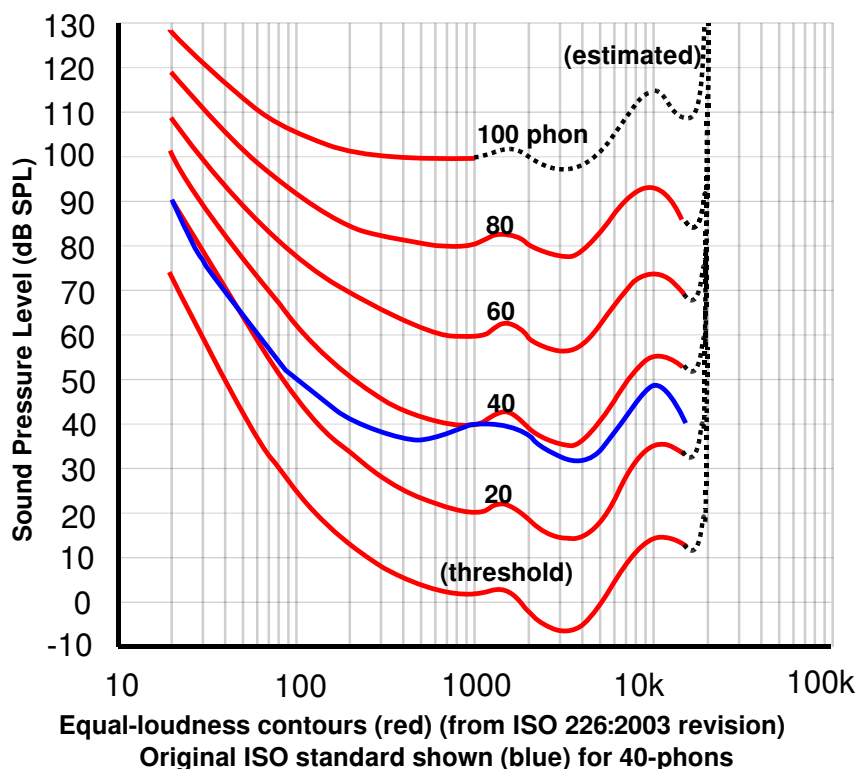


图 1.4: 听觉感知的响度级

1.3 语音的基本特征

和人类生活关系最密切的大概就是语音了。语音是人的发声器官激发空气振动而产生的声波。在地球的生物中, 语言可能是人类独有的能力。语言的主要目的是传递信息。根据 Shannon⁴ 的信息理论, 一系列离散符号表示的信息容量可以量化为比特, 衡量信息传输的速度用 bits/s (bps) 表示。在各种有关语音信息的处理过程中, 语音信号 (声波, 模拟信号) 会经过量化、编码、传输、记录、处理等不同的过程, 最终还原成声波模拟信号并被我们的听觉系统所感知。

1.3.1 语音的传输过程

图1.5是一个从声波发出到接收的一个完整的过程。

说话人一方将文字符号或大脑中思考的信息转换成音素, 由神经系统激励发声器官产生声波输出。在大多数音频处理设备中, 声波经传声器拾取转换成电波。系统的另一边通过扬声器或耳机将电波还原成声波 (数字系统还应经过 A/D 和 D/A 转换)。声波作用在受话人耳膜, 刺激基底膜神经元。基底膜如同一个谱分析仪, 不同频率对不同的感知位置刺激反应不同, 神经系统将其解码, 提取出音素、词汇和语句, 最终完成语言理解。

⁴ Claude Elwood Shannon (1916.4.30 – 2001.2.24), 美国数学家、电气工程师、密码学家, 信息论的创始人。

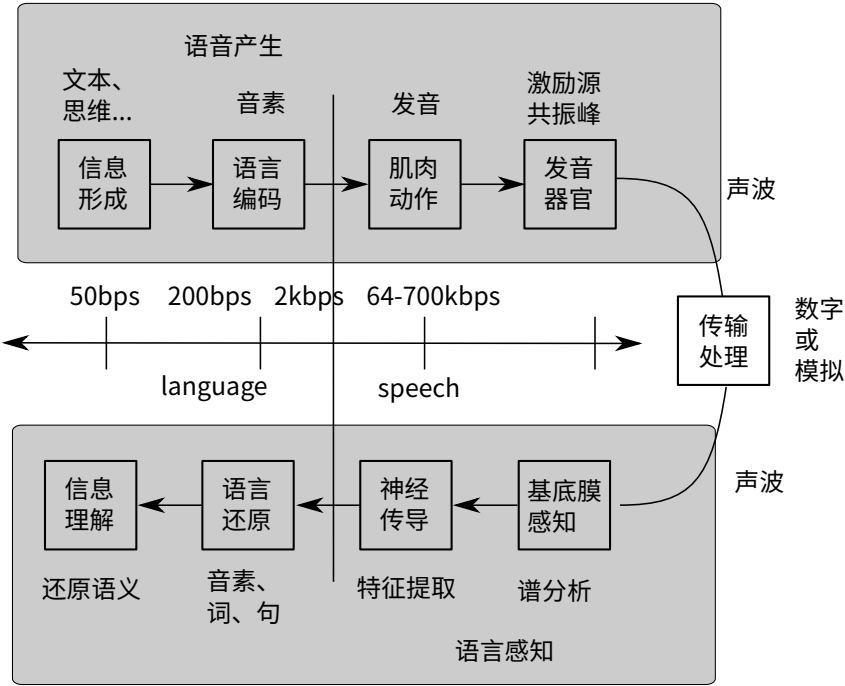


图 1.5: 语音传输过程

1.4 常用音频仪器

1.4.1 音频设备测量

传声器、耳机、扬声器是常见的音频设备. 设备出厂时都需要对它们进行定标.

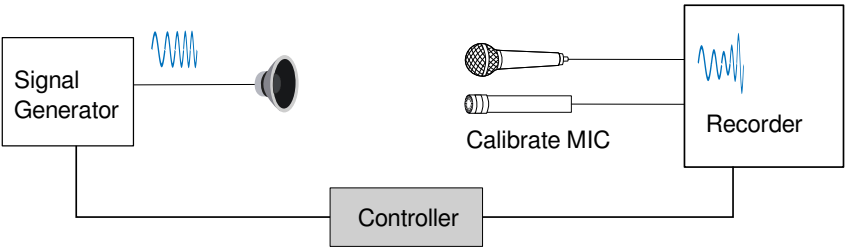


图 1.6: 传声器测量 (对比法/置换法)

作为校准的传声器通常是电容传声器. 传声器电容与隔膜和背板之间的距离成反比. 当暴露有声场中时, 通过测量传声器的电容即可得到位移 (或声压).

测量时, 通常由信号源产生幅度一致的单频或扫频信号. 接收端与信号源连动, 驱动记录仪存储或绘制传声器的幅频响应.

电声/声电传感器设备工作时, 很容易受到“多径效应”的影响⁵. 系统传递函数会对被测对象产生很大的影响. 在数字系统广泛应用之前, 通常这样测量都在消声室 (anechoic chamber) 中

⁵ 多径效应是指信号通过两个或两个以上路径到达接收端的情况. 此时接收端收到的信号为发射端信号与传输路径的卷积.

进行, 使传递函数成为一个 $\delta(t)$ 函数. 大多数情况下, 传声器的频响都可以做得比扬声器单元更宽和更平坦. 数字测量手段主要集中在针对扬声器测量上.

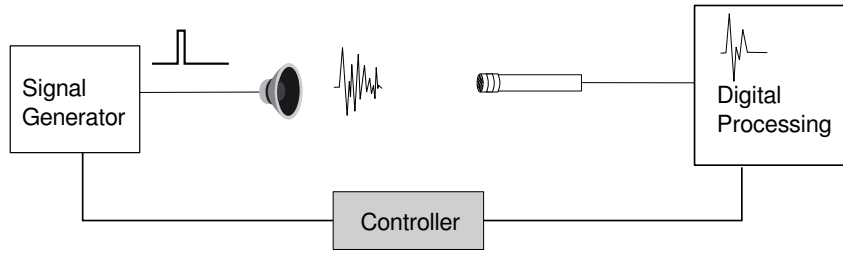


图 1.7: 扬声器非消声室测量

早期一种较为典型的数字测量技术是短时截断法, 即, 给扬声器一个短促的脉冲, 截取接收端信号一个较短的时间响应, 以确保没有反射声的影响, 对其做频谱分析. 我们知道, 冲激函数或阶跃函数包含了所有的频率成分, 从而也就能算出扬声器的频率响应.

1.4.2 声音采集和还原系统

传声器将声压转换成电压, 经 A/D 转换器采集成数字信号. 传声器的灵敏度决定了电压与声压的关系. 传声器灵敏度表示为 1kHz 信号产生 94dB (或 1Pa) 声压级的电压, 它表明在给定 SPL 时输出信号是多少伏, 习惯上以 dBV 表示.

$$Sensitivity_{dBV} = 20 \lg \left(\frac{Sensitivity_{mV/Pa}}{dBV_{ref}} \right)$$

参考电压 dBV_{ref} 为 1V/Pa。

一个标称 “94dB/50mV” 灵敏度的传感器意味着 1Pa 的声压输出 50mV。A/D 转换后的数值量可据此定标。同样的情况也出现在 D/A 转换后的模拟声音还原过程。

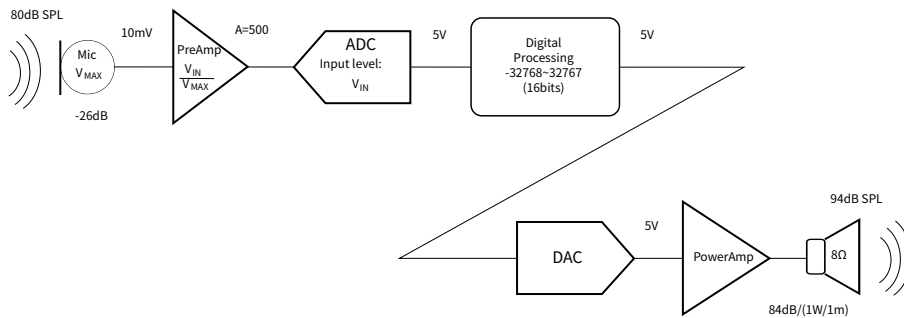


图 1.8: 模拟量声压在数字处理过程中的定标

用标准的声级计测量扬声器的声音输出, 为了保持一致性, 声级计置于距离扬声器 1 米处测量声压, 放大器提供 1 瓦的电功率. 因此, 扬声器灵敏度通常表示为 “84dB(1W/1m)” 这种形式.

对于给定的扬声器, 当功率增加一倍时, 距离 1m 处的声压级增加 3dB; 功率增加 10 倍, 声压级增加 10dB (不是 20dB!).

由于已经将声音系统作为线性系统看待, 因此大多数数字系统并不关心模拟物理量的绝对值.

2

语音参数分析



2.1 语音产生的生理模型

2.1.1 音位

音位 (Phoneme) 是语音学中的概念, 它是人类语言中具有区别意义的最小声音单位。一个词由一至数个音节组成, 一个音节由一至数个“音段”(元音、辅音等) 组成。音位被定义为能区分语义的“音段”。如果两个声音所代表是同一个词汇、同样的意义, 即使不同的“音段”也被视为同一个音位。音位与语言相关, 每一种语言都有自己的一组音位, 他们构成了这个语言的语音系统。音位可用来研究某个特定语言中如何将音组合成词。¹

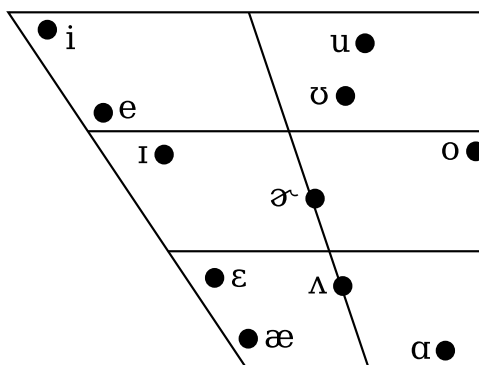


图 2.1: 音位图

音位并不是一个实际的语音。一个音位可以有数种不同的发音, 但是语言使用者心理上认定这几种发音是等同的, 甚至可能不会察觉语音上有所不同。确定是否为音位的准则有: 一、辨义功能; 二、互补分布; 三、音感差异。

不同语言中音位不同。例如, 在汉语的一些方言中, 鼻音和舌尖音不分、平翘舌不分, 是因为这些音在这种方言里是同音位的, 并不造成语言理解上的困难。

¹音位有时被译为“音素”, 然而音素一词在中文里的用法较为混乱, 不一定都是指音位。

2.1.2 发音器官

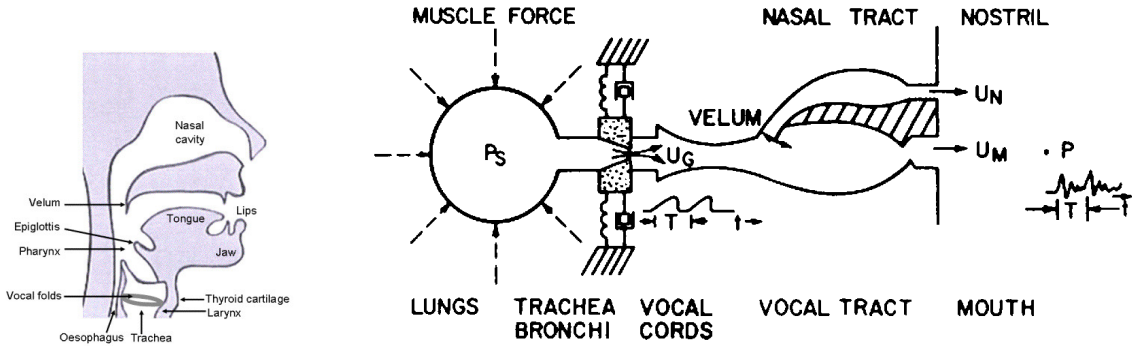


图 2.2: 发音器官及其等效物理模型

图2.2是语音研究领域一种简化的“激励源-滤波器”模型。肺部的气压冲击声带引起声带振动, 作为激励源, 产生宽带信号(脉冲)。声带振动可近似认为是一个周期脉冲, 它包含较丰富的频谱。信号通过口腔及鼻腔的传播途经, 相当于经过一个管道滤波器, 最终声音从管口输出。

在一些语音模型中, 典型的如 LPC (Linear Predictive Coding 线性预测编码器), 管道的截面积参数就可以通过一组系数来表示。

2.2 语音频谱特征

语音信号是时变信号, 所以频谱也是随时间变化的。需要说明, 由于目前平稳随机信号理论相对成熟, 绝大多数语音分析也是基于平稳随机信号的理论进行处理的。将非平稳随机信号转换成平稳随机信号处理的形式, 常采用短时频谱分析技术 (STFT—Short Time Fourier Transform):

$$X(t, f) = \int_{-\infty}^{\infty} w(t - \tau)x(\tau)e^{-j2\pi f\tau} d\tau$$

其离散形式是:

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} w(n - m)x(m)e^{-jm\omega}$$

式中的 $w()$ 是窗函数, 其长度取决于信号的平稳特性。对于元音信号, 可以认为在较长时间 (大概是 20ms) 频谱特性是稳定的, 而对于部分辅音 (如阻塞音、爆破音), 分析窗可能需要缩短至 5ms。通过 STFT 分析的短时频谱只反映语音信号静态频谱特性。对于一段有意义的语音, 可将与时间相关的傅立叶分析结果展示为三维频谱图, 以时间为横坐标、频率为纵坐标。这种用灰度表示的频谱图称为 sonagraph 或 spectrograph, 它表示语音频谱随时间变化的图形。有经验的语音学家可以根据语谱图“看”出其中的语言信息 (如音素、发音人), 因此这种图又被称为“声纹” (对应指纹)。最初的语音识别和说话人识别即是由由此开始。

线性系统的短时频响可表示为如下的函数形式:

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{1 - \sum_{k=1}^N a_k z^{-k}} = \frac{b_0 \prod_{k=1}^M (1 - d_k z^{-1})}{\prod_{k=1}^N (1 - c_k z^{-1})} \quad (2.1)$$

滤波器系数 a_k 、 b_k 的变化率约是 50–100 次/秒, 极点 (c_k) 靠近单位圆, 它们形成了语音的共振峰。零点 (d_k) 用来模拟鼻音和摩擦音。

2.2.1 音调

大多数乐音和语音中的浊音 (元音及浊辅音) 都有周期性, 这个周期性参数被称为音调 (pitch)。音调是一个与信号基频相关的主观属性, 它与线性频率的关系使用与 Mel 相同的公式 (1.6), 它与频率近似成对数关系。该公式经过校正, 1000Hz 对应 1000 mels。然而, 在科学研究中, 音调更多的是直接用频率单位表示。

汉语是一种典型的有调语言, 其调性的变化就是通过音调反映的。在一些场合, 这个“调”又被称作“音高”。图2.3是汉语四声“山明水秀”的窄带语图, 其中起伏变化的暗色条纹就是基频。

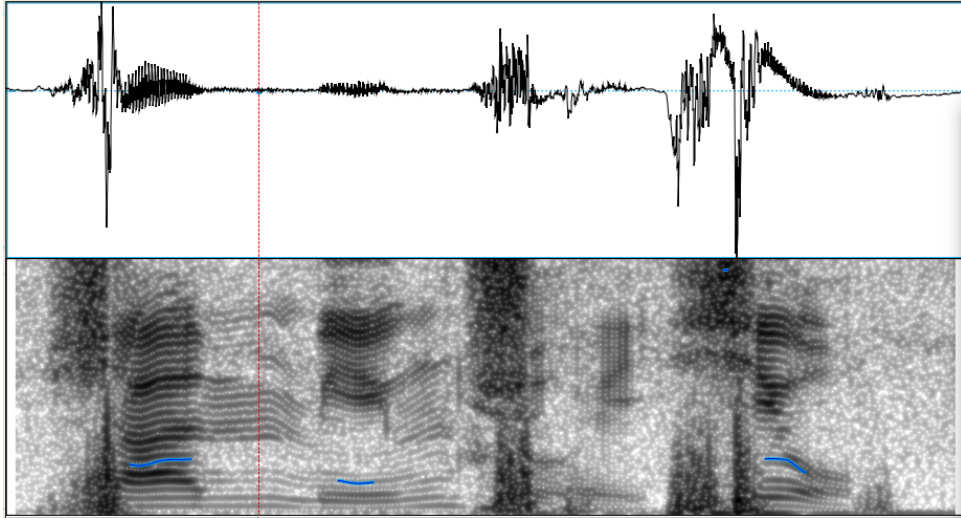


图 2.3: 四声“山明水秀”的窄带语图

在发音时, 声带的振动频率决定音高。正常发音时, 男声的音高在 80Hz 到 300Hz 范围附近, 女声的音高在 200Hz 到 400Hz 附近, 而专业歌唱演员的音域要宽得多。表2.1是不同声部歌唱演员的音调范围。

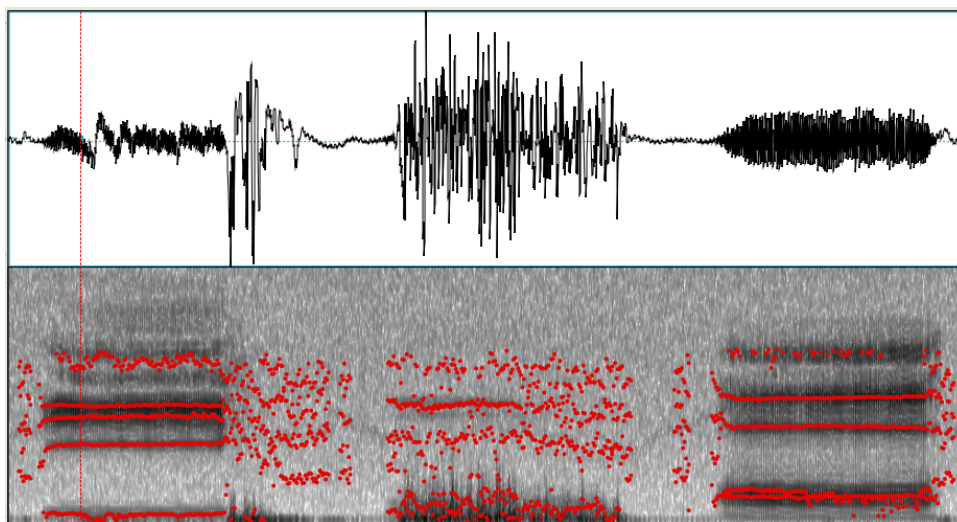
2.2.2 共振峰

共振峰是频谱中能量相对集中的一些区域。在声学中, 共振峰用来描述声源内部的共振, 对乐器而言, 指的是共鸣箱内的共振。在语音研究领域, 声道的特性形成了不同声音的共振峰。反

表 2.1: 不同声部歌唱演员的音域

声部	音程	频率范围 (Hz)
女高音	C4(中央 C) 到 C6	261—1046
女中音	A3 到 A5	220—880
女低音	F3 到 E5	174—659
男高音	B2 到 A4	123—440
男中音	G2 到 F4	98—349
男低音	E2 到 E4	82—329

过来说, 共振峰反映了声道的物理特征。图2.4是元音 $[i, u, a]$ 的宽带语图, 能量集中的部分就是它们对应的共振峰。表2.2是英语主要元音的前两个共振峰的频率。

图 2.4: 元音 $[i, u, a]$ 的宽带语图

2.3 短时分析

图2.5是常用的语音产生激励模型, 有声 (voiced) 段来自声带的周期激励, 无声段 (unvoiced) 则为随机噪声序列。在有音段, 声门激励脉冲与声道系统卷积。声道系统缓慢变化, 在较短时间内, 其参数 (频响) 保持不变。该系统常常被描述为一个全极点系统:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.2)$$

其差分形式是:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Ge(n) \quad (2.3)$$

表 2.2: 常见元音的共振峰平均值

元音	共振峰 F1(Hz)	共振峰 F2 (Hz)
iy (feel)	300	2300
ih (fill)	360	2100
ae (gas)	750	1750
aa (father)	680	1100
ah (cut)	720	1240
ao (dog)	600	900
ax (comply)	720	1240
eh (pet)	570	1970
er (turn)	580	1380
ow (tone)	600	900
uh (good)	380	950
uw (tool)	300	940

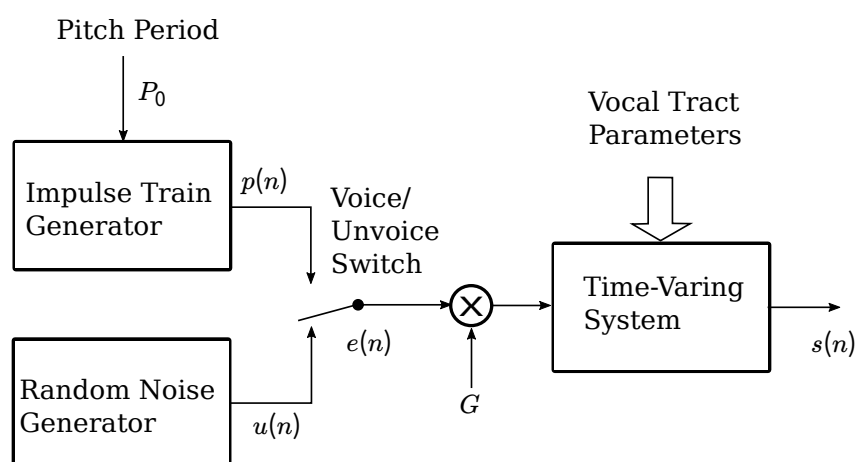


图 2.5: 语音信号 Voiced/Unvoiced 模型

由于参数 a_k 、 G 是随时间变化的, 考虑非时变系统的分析基础, 将语音分段处理, 每一段称为一帧 (frame), 此为短时分析的基础。

2.3.1 短时语音参数

常用的语音短时参数有短时能量和短时过零率。它们计算简单, 可以比较方便地表现语音的一些特征。

短时能量

短时能量定义为

$$\begin{aligned} E &= \sum_{m=-\infty}^{\infty} (x(m)w(n-m))^2 \\ &= \sum_{m=-\infty}^{\infty} x^2(m)w^2(m) \end{aligned}$$

短时能量可作为有声/无声的重要判据。

短时过零率

短时过零率定义为

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} |\text{sgn}(x(m)) - \text{sgn}(x(m-1))| w(n-m) \quad (2.4)$$

这里 $\text{sgn}()$ 是符号函数, 它被定义为

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (2.5)$$

当序列 $x(n)$ 频繁改变正负号, Z_n 就是一个较大的值。它可以用来判断信号的高频成份: 清辅音的过零率高于元音和浊音。图2.6是一段语音的短时能量和过零率。

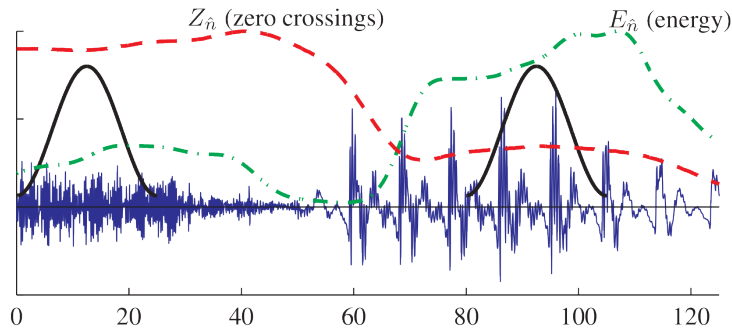


图 2.6: 一段语音的短时能量和短时过零率

2.3.2 STFT

短时分析是语音分析的重要手段。短时能量和短时过零率只能部分分析信号的频率特性。完整频率分析需要用到短时傅立叶变换, 它被定义为

$$X(n, e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m} \quad (2.6)$$

短时傅立叶变换包含两个参数: 时间参数 n 和频率参数 ω , 它克服了傅立叶变换的时间局限。

将式 (2.6) 写成卷积形式

$$X(n, e^{j\omega}) = (x(n)e^{-j\omega n}) * w(n) = (x(n) * w(n)) e^{-j\omega n} \quad (2.7)$$

从滤波器的角度看, 大多数窗函数相当于一个低通滤波, 因此 $X(n, e^{j\omega})$ 在给定 ω 轴上是随 n 缓慢变化的信号。

式 (2.6) 在频率上离散化, 得到

$$X(r, k) = \sum_{m=rR-L+1}^{rR} x(m)w(rR-m)e^{-j(2\pi k/N)m}, k = 0, 1, \dots, N-1 \quad (2.8)$$

在频率轴上均匀分布 N 点的频谱。上式换一种表示

$$X(r, k) = \hat{X}(r, k)e^{-j(2\pi k/N)rR} \quad (2.9)$$

其中

$$\hat{X}(r, k) = \sum_{m=0}^{L-1} x(rR-m)w(m)e^{j(2\pi k/N)m} \quad (2.10)$$

是 $x(rR-m)w(m)$ (即在第 r 帧加窗) 的 DFT。 L 是窗的长度。短时离散傅立叶计算流程如下:

1. 对时间序列 $x(n)$ 在 r 位置上加窗, 得 $x_{rR}(m) = x(rR-m)w(m)$, 长度为 L ;
2. 计算 $x_{rR}(m)$ 序列共轭 DFT (可利用 FFT 算法)
3. 乘上因子 $e^{-j(2\pi k/N)rR}$, 实际上, 在大多数分析场合是不必要的, $\hat{X}(r, k)$ 对应一帧的语谱 (见图2.3和2.4):

$$S(r, k) = 20 \lg |\hat{X}(r, k)| = 20 \lg |X(r, k)| \quad (2.11)$$

4. 时间序列前移 R , 计算下一帧。

帧移 R 和 N 由下面的因素决定:

- $R \leq L/(2C)$, 其中常数 C 由窗函数带宽决定, 对于 Hamming 窗, $C = 2$, 对于矩形窗 $C = 1$ 。此项要求保证在时间上采样率高于带宽的 2 倍, 频率上不产生混叠。
- $N \geq L$, 避免时间混叠。

2.4 线性预测编码

由语音的全极点模型 (图2.5) 的差分形式 (公式2.12), 容易想到, 语音信号 $s(n)$ 可以表示为前若干样点的线性叠加, 即:²

$$\hat{s}(n) = \sum_{k=1}^P \alpha_k s(n-k) \quad (2.12)$$

²在一些教材中, 将系数 α_k 表示为负数。公式2.13 被写成 $e(n) = s(n) + \sum_{k=1}^P \alpha_k s(n-k) = \sum_{k=0}^P s(n-k)$, 其中 $\alpha_0 = 1$, 形式看上去更紧凑。

预测误差

$$e(n) = s(n) - \hat{s}(n) \quad (2.13)$$

这里的问题变成了：从语音信号中获得一组预测系数 α_k ，用于估计声道参数。定义均方误差

$$\begin{aligned} En &= E\{e^2(n)\} \\ &= E\left\{s(n) - \sum_{k=1}^P \alpha_k s(n-k)\right\}^2 \\ &= \frac{1}{N} \sum_n \left(s(n) - \sum_{k=1}^P \alpha_k\right)^2 \end{aligned} \quad (2.14)$$

通过 $\partial En / \partial \alpha_i = 0$ ，求解一组 $\{\alpha_k\}$ ，使误差最小化，即得到如下一组线性方程

$$\sum_{k=1}^P \alpha_k \sum_m \{s(m-i)s(m-k)\} = \sum_m \{s(m-i)s(m)\}, 1 \leq i \leq P \quad (2.15)$$

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(P-1) \\ \phi(1) & \phi(0) & \cdots & \phi(P-2) \\ \cdots & \cdots & \cdots & \cdots \\ \phi(P-1) & \phi(P-2) & \cdots & \phi(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_P \end{bmatrix} = \begin{bmatrix} \phi(1) \\ \phi(2) \\ \cdots \\ \phi(P) \end{bmatrix} \quad (2.16)$$

或写成矩阵形式：

$$\Phi \alpha = \phi \quad (2.17)$$

自相关系数矩阵 Φ 是正定对角矩阵。 ϕ 为自相关函数。求解该方程采用 Levison-Durbin 算法。

3

求解该方程的过程中同时得到一组 k 参数，从算法过程可以看出，它满足如下方程：

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \quad (2.18)$$

如果将声道看作一个管道， k_i 相当于传输系数。管道中的声波传输系数与截面积相关，因此系数 k_i 可视作是与声道截面相关的一组参数。

图2.7是语音 a 其中一段的频谱，图2.8是 LPC 系数和对应的声道的特性。注意到，声道特性频谱与短窗对应的短时傅立叶谱非常相似。

2.5 同态信号处理

考虑信号 $s(n)$ 通过一个系统 $h(n)$ ，时域的卷积关系，反映在频域则是相乘。对于语音的激励模型， $s(n)$ 对应激励源， $h(n)$ 对应声道。联想到对数运算可以将乘法变成加法，如果对数域能将 $s(n)$ 与 $h(n)$ 分离，就可以通过这种手段实现特征提取。

³J.Marhonl, *Linear Prediction: A Tutorial Review.*, Proc.IEEE Vol.63,pp561-580, 1975

Algorithm 1 Levison-Durbin 算法

```

 $E^{(0)} = \phi(0), \alpha_1^{(0)} = 1$ 
for  $i = 1, 2, \dots, P$  do
     $k_i = \left( \phi(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} \phi(i-j) \right) / E^{(i-1)}$ 
     $\alpha_i^{(i)} = k_i$ 
    if  $i > 1$  then
        for  $j = 1, 2, \dots, i-1$  do
             $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$ 
        end for
    end if
     $E^{(i)} = (1 - k_i^2) E^{(i-1)}$ 
end for
 $\alpha_j = \alpha_j^{(P)}, j = 1, 2, \dots, P$ 

```

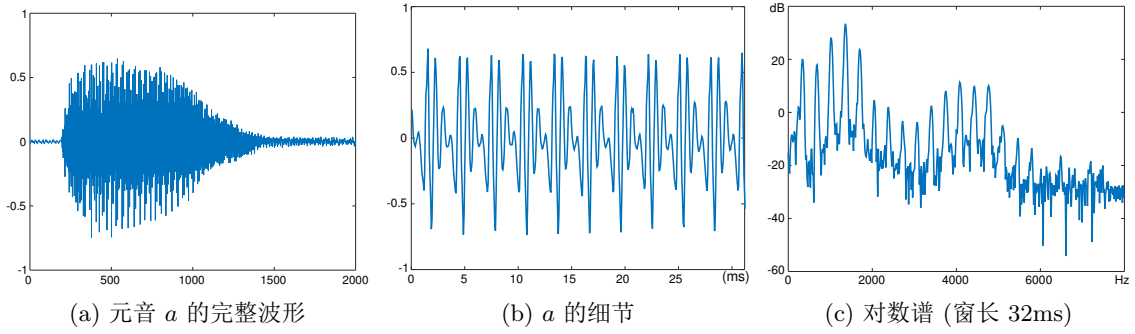
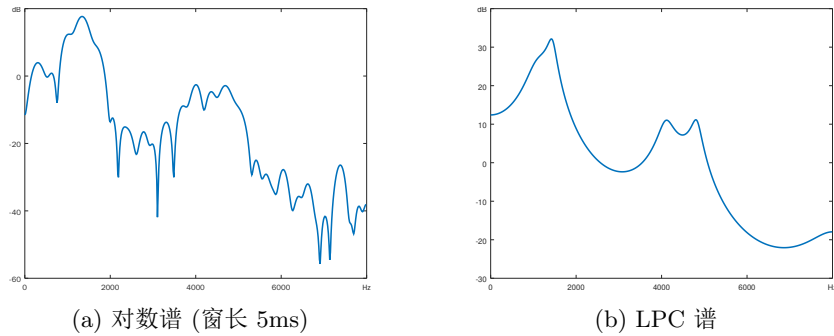
图 2.7: 元音 a 波形和频谱

图 2.8: 短窗频谱和 LPC 对数谱

图2.7(c) 中的基频在频域对数曲线上显示明显的周期型。

这里引入倒谱的概念 (cepstrum, spectrum 倒读)。离散时间信号的倒谱定义为

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega \quad (2.19)$$

去掉上式的绝对值符号, 保留复数信息, 按类似的变换, 称为复倒谱。复倒谱处理比较复杂, 因为相位可能延伸到 $(-\pi, +\pi)$ 范围之外。 $c(n)$ 的下标为“倒频域”(quefrequency, frequency 倒读), 它具有时间的量纲。

短时倒谱可使用 IFFT 算法。在语音处理中, 倒谱常作为基音检测的工具。由于声道和激励在倒谱中可以很好地分离, 所以很自然地考虑使用倒谱分析技术来估计声道系统。图2.9是图2.7的语音在倒谱域上进行分离的处理结果。

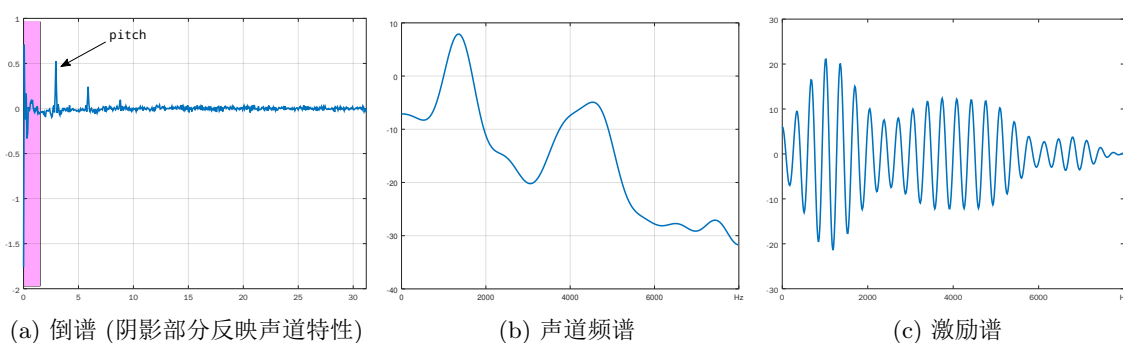


图 2.9: 倒谱分离声道和激励

图2.10是“欢迎”一词随时间变化的短时频谱和倒谱轨迹。

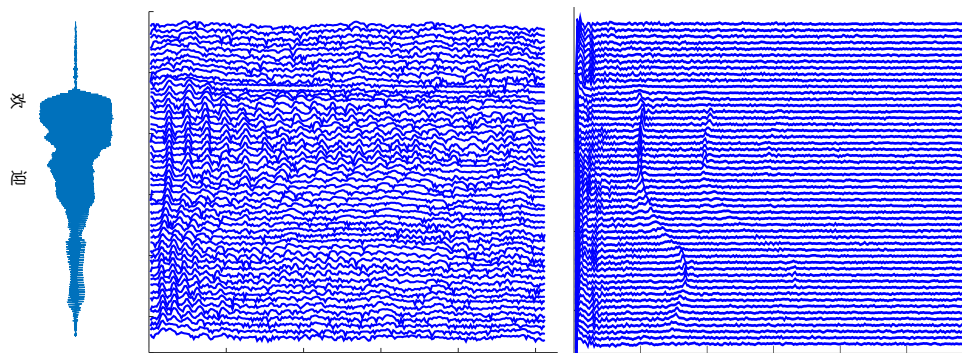


图 2.10: “Huan Ying” 波形、spectrograph、cepstrum

3

数字滤波系统



根据数字信号处理理论, 信号通过一个 (滤波) 系统, 相当于频域的乘积。但在实际工程实现中, 大量的处理仍在时域完成。

数字信号处理的实现方法一般有以下几种:

- 在通用计算机上用软件实现, 可以使用专业软件如 Matlab 或通过其他高级编程语音。偏重于算法验证, 不考虑实时实现。
- 在同样计算机上增加一些专用高速数字信号处理集成电路, 可实现一定程度的实时应用, 但由于依赖通用计算机, 不适合商业化产品。
- 采用可编程嵌入式处理器 (特别是 DSP 芯片) 实现实时系统, 可商业化。
- 使用 FPGA, 可满足实时性要求, 且成本比商用 DSP 处理器更低, 但灵活性不如后者。

滤波是最常规的处理内容。广义上说, 所有系统都可以称做滤波器。这里仅讨论具有预期频率特性的系统, 即经典信号与系统中的频率滤波器。

3.1 有限冲激响应滤波

3.1.1 理想滤波器模型

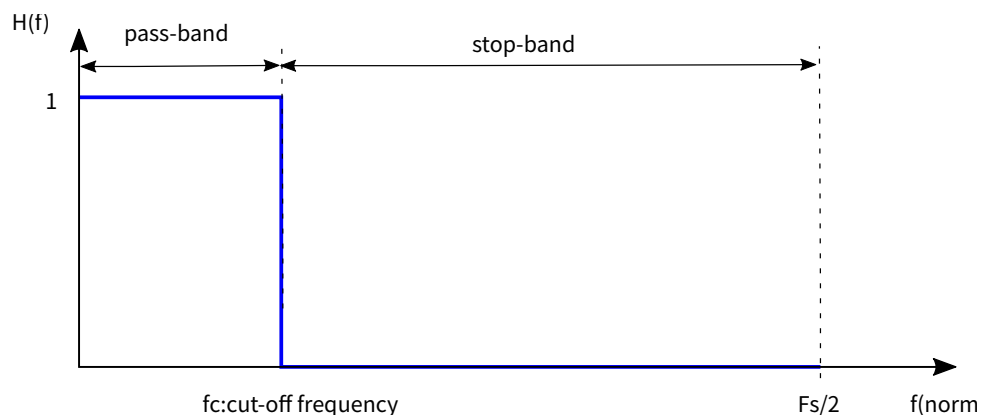
由于理想滤波器的不可实现, 实际实现的滤波器频响在通带或止带会有一些起伏, 并且过渡带的宽度也不为零。为了让滤波器更接近于理想, 需要综合考虑其中的一些指标。

3.1.2 冲激响应计算

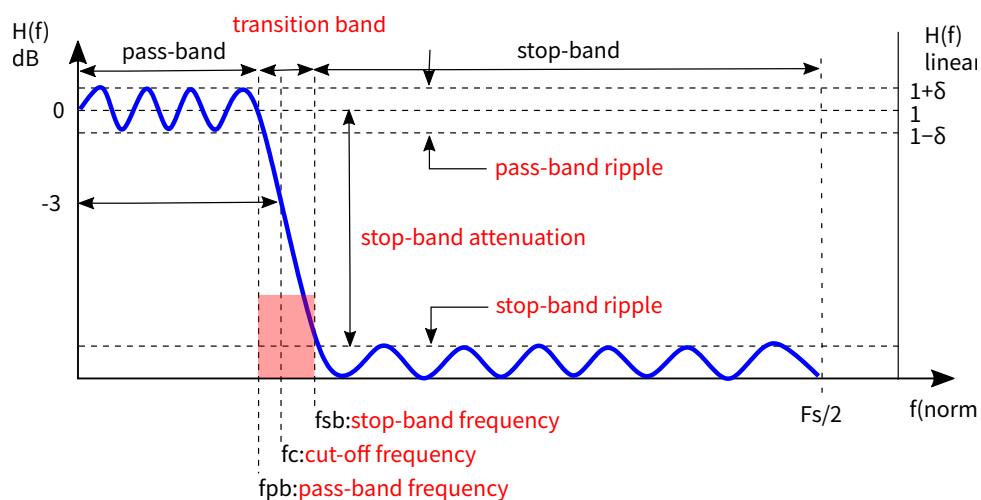
图3.2是有限冲激响应滤波器的横向结构。信号输入 $x(n)$ 、滤波器参数 b_k 和系统输出 $y(n)$ 之间的关系是:

$$y(n) = \sum_{k=0}^{N-1} b_k x(n-k) \quad (3.1)$$

将信号 $x(n)$ 用冲激函数 $\delta(n)$ (Dirac 函数) 替代, 可以看出, 滤波器参数 b_k 就是冲激响应, 其中 N 是滤波器阶数。由于滤波器长度有限, 因此 FIR 滤波器总是稳定的。



(a) 理想低通滤波器



(b) 实际滤波器指标

图 3.1: 理想滤波器与实际滤波器

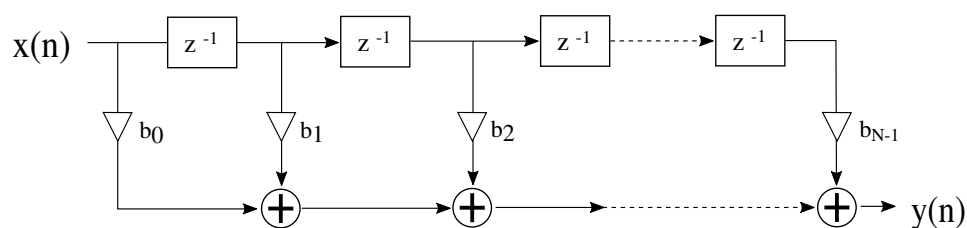


图 3.2: 有限冲激响应滤波器横向模型

工程应用中, 多数情况需要计算出满足频响特性 $H(e^{j\omega})$ 的滤波器系数。计算方法很多, 比较重要的有:

- 窗函数法
- 频率取样法
- Parks-McClellan 迭代法

注意, 不能将 $H(e^{j\omega})$ 直接通过 IDFT 变换得到的系数作为滤波器系数, 因为离散傅立叶变

换只在离散频率点才有定义, 在频域非采样点上没有定义, IDFT 并不能默认地用相邻数据作为内插。但 IDFT 在 FIR 设计和滤波过程中仍然是非常有意义的。

窗函数设计法

这里以理想低通滤波器为例, 了解窗函数法的设计过程。设滤波器数字域频响为:

$$H(e^{j\omega}) = \begin{cases} 1, & \omega < |\omega_c|, \\ 0, & \text{else} \end{cases} \quad (3.2)$$

该理想滤波器的系数

$$\begin{aligned} h_d(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega \\ &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} 1 \cdot e^{j\omega n} d\omega \\ &= \begin{cases} \frac{\sin(n\omega_c)}{n\pi}, & \text{for } n \neq 0 \\ \frac{\omega_c}{\pi}, & \text{for } n = 0 \end{cases} \end{aligned} \quad (3.3)$$

这是一个无穷级数, 为了物理可实现, 需要做两件事: 一、截取其中的一段, 保留有限长序列; 二、在时间轴上平移, 以使其为因果系统。

窗函数截取的阶数越多 (N 越大), FIR 滤波器的过渡带越接近理想值; 窗函数的形式还影响 FIR 滤波器的通带纹波和止带衰减。表3.1是不同窗函数对滤波器设计结果的影响。

表 3.1: 窗函数对 FIR 性能的影响

窗函数	归一化过渡带宽 Width($\Delta f(\text{Hz})$)	通带 纹波 (dB)	止带 衰减 (dB)
矩形窗	$0.9/N$	0.7461	21
Hann 窗	$3.1/N$	0.0546	44
Hamming 窗	$3.3/N$	0.0194	53
Blackman 窗	$5.5/N$	0.0017	74
Kaiser 窗	$2.93/N \rightarrow \beta = 4.54$	0.0274	50
	$5.71/N \rightarrow \beta = 8.96$	0.000275	90

表中可以看出, 随着阶数的增加, 过渡带变窄, 也就是在频域上过渡带更加陡峭。矩形窗的主瓣最窄, 但是旁瓣的幅度最大。在阶数相同的条件下, 矩形窗的过渡带是最窄的, 除此以外, 矩形窗就没其他优势了, 它的纹波最大, 且止带只能衰减 21dB。

式 (3.3) 中, 经过平移, 实际滤波器冲激响应是

$$h(n) = \sum_{k=0}^{N-1} w(k) h_d(n-k), n = 0, 1, \dots, N-1 \quad (3.4)$$

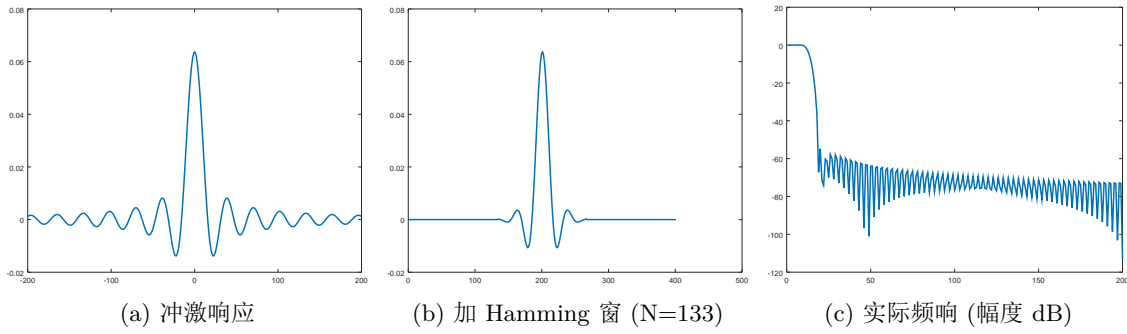


图 3.3: FIR 窗函数设计过程和结果

图3.3是窗函数设计滤波器的结果。根据傅立叶变换原理,系统的时间平移对应频率响应相位加一个常数。如果窗函数是对称的,式(3.4)是一个线性相位系统。

FIR 滤波器最大的优势是,它可以构成线性相位。构成线性相位 FIR 滤波器需要满足对称性条件。即当 $h(k) = \pm h(N-1-k)$, 且 N 为偶数时,

$$\begin{aligned}
 H(e^{j\omega}) &= \sum_{k=0}^{N-1} h(k)e^{-j\omega k} \\
 &= e^{-j\omega \frac{N-1}{2}} \sum_k^{[N/2]-1} h(k) \left(e^{j\omega(\frac{N-1}{2}-k)} \pm e^{-j\omega(\frac{N-1}{2}-k)} \right)
 \end{aligned} \quad (3.5)$$

上式括号中可以合并为 $\sin()$ 或 $\cos()$, 求和号内的结果为实数, 相位仅由求和符号外部分决定。对于 N 为奇数的情况可以得出同样的结论。

频域取样法

对理想滤波器 $H_d(e^{j\omega})$ 在频域取样

$$H_d(k) = H_d(e^{j\omega})|_{\omega_k = \frac{2\pi}{N}k} = H_d(e^{j\frac{2\pi}{N}k}) \quad (3.6)$$

再对 $H_d(k)$ 做 IDFT, 得到 N 点单位取样序列

$$h(n) = \frac{1}{N} \sum_{k=0}^{N-1} H_d(k) e^{j\frac{2\pi}{N}nk}, n = 0, 1, \dots, N-1 \quad (3.7)$$

滤波器的频率响应

$$H(e^{j\omega}) = \frac{1}{N} \sum_{k=0}^{N-1} H_d(k) \frac{1 - e^{-j\omega N}}{1 - e^{j\frac{2\pi}{N}k} e^{-j\omega}} \quad (3.8)$$

通过频率取样法设计的 FIR 滤波器, 其频响在频率取样点上严格等于期望值, 不在频率取样点上的频响由内插函数的插值决定。

频域取样法利用快速傅立叶变换工具, 将频率特性反变换到时间域, 得到冲激响应。为了让滤波器的频率特性更优, 可以增加频率取样的密度, 并在频率跳跃点增加一些缓变的过渡, 再通

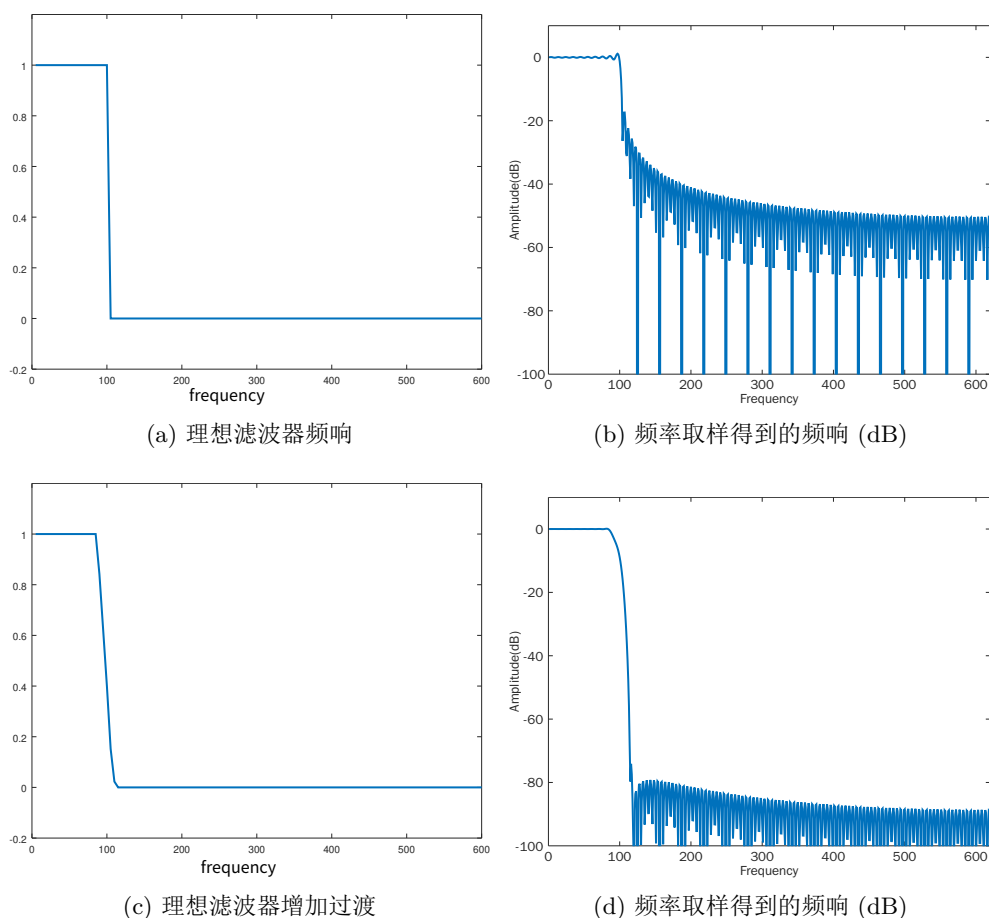


图 3.4: 频域取样法设计过程和结果

过窗函数截取、平移, 计算出实际 FIR 滤波器系数。图3.4是两种不同的取样方法设计的滤波器特性。可以看出, 通过增补过渡带, 通带纹波和止带衰减都得到了改善。

频域取样法滤波器设计步骤:

1. 根据理想滤波器 $H_d(k)$ 指标, 在频率上等间隔取样; 为实现实序列冲激响应, 根据 DFT/IDFT 性质, 应将 $H_d(k)$ 设置为对称。
2. 对称性要求 $H_d(N/2) = 0$, 故高通、带阻不能是偶数阶次。
3. 利用 IDFT 计算冲激响应 $h(n) = \text{IDFT}\{H_d(k)\}$;

为减小纹波, 适当增加过渡带或增加阶次。

Parks-McClellan 算法

Parks-McClellan algorithm 算法, 又名切比雪夫逼近、Remez-exchange 算法、最小化最大误差算法, 由 James McClellan 和 Thomas Parks 于 1972 年提出。该算法的主要思想是利用迭代的方式最小化滤波器在通带和止带的最大误差。

3.1.3 滤波算法

FIR 滤波快速算法

在串行计算机上, 滤波算法可以按式3.1进行。通常 FIR 滤波器阶数比较大, 一次滤波都可能成百上千次的乘加运算, 相当耗时。工程上, 为了减少计算量, 利用快速傅立叶变换是一个有效的手段。

设有 1000 阶的 FIR 滤波器 $h(n)$, 处理 1000 个点的数据 $x(n)$ 就需要 1 百万次运算。通过线性卷积算法, 将 1000 阶 FIR 滤波和 1000 个采样点补零, 各自扩展到 2048 点, 用 FFT 计算出二者的频域值, 在频域相乘, 反变换到时域。这期间发生了两次 FFT 和一次 IFFT, 以及 N 次频域乘法。FFT 和 IFFT 算法的复杂度都是 $\frac{N}{2} \log_2 N$, 当 $N = 2048$ 时, 三次 FFT 总的乘法次数是 $2048/2 * \log_2(2048) * 3$, 大概是三万多乘法, 即使考虑到复数运算以及其他一些开销, 计算量也远远小于 1 百万次。并且, 由于滤波器系数相对比较稳定, 并非每帧数据都需要对 $h(n)$ 做 FFT。

设信号 $x(n)$ 、FIR 冲激响应 $h(n)$, 通过循环卷积计算线性卷积, 步骤如下:

1. 将序列 $x(n)$ 和 $h(n)$ 补零延长, 使其长度 $L \geq M + M - 1$ 。如不考虑系统延时, 取序列 $x(n)$ 的长度 N 与滤波器阶数 M 接近, 可使 FFT 效率更高。如果采用基-2 FFT, L 还应是 2 的整数次幂。
2. 计算补零后的 $x(n)$ 和 $h(n)$ 的 FFT (长度 L), 在频率域计算它们的积 $Y(k) = H(k) \cdot X(k)$ 。如果滤波器参数没有更新, $H(k)$ 不需要每帧都计算。
3. 计算 $Y(k)$ 的 IFFT, 得到 $y(n)$ 。
4. 对于得到的 $y(n)$, 主要有两种处理方法: 重叠相加法和重叠保留法。这里仅介绍重叠相加法, 即, 取 $y(n)$ 的前 N 个点, 作为当前帧的输出 (假设此时是信号的开始, 之前的信号为 0); 将剩余部分保留, 用来叠加到下一帧的输出。下一帧的 $y(n)$ 的前 N 个点与保留值叠加, 作为本帧的输出, 后 N 个点继续保留以备后用。依次类推。

滤波结构选取

图3.2是一种较常用的结构, 但并不是唯一可用的结构。将滤波器变形, 形成级联结构:

$$\begin{aligned} H(z) &= \sum_{k=0}^{N-1} b_k z^{-k} \\ &= b_0 \prod_{k=1}^M (1 + b_{k,1} z^{-1} + b_{k,2} z^{-2}) \end{aligned} \quad (3.9)$$

图3.5是级联算法结构。在这个结构中, 每一级对应滤波器的一对零点 (当 N 为奇数时还应有一项一阶节)。这种算法的优点是可以独立控制零点, 缺点是参数变多, 增加计算量。

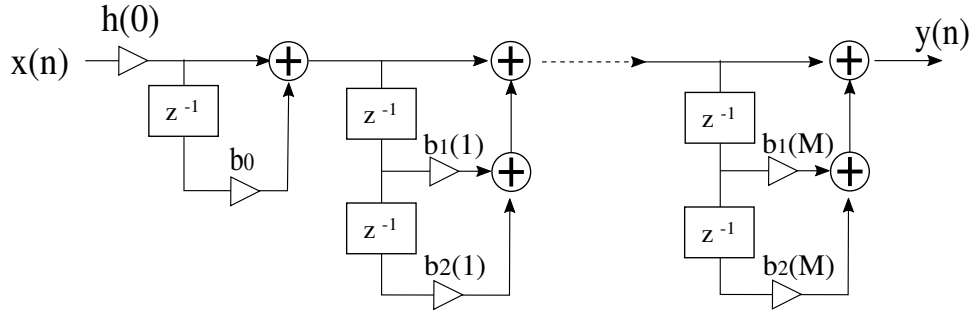


图 3.5: FIR 级联结构

3.2 无限冲激响应滤波

IIR 滤波器由于采用反馈 (或设置极点), 在满足同等幅频响应的前提下, 可利用较少的系数得到更高的计算效率。但由于极点的存在, 系数计算不当, 可导致滤波器的不稳定。幅频响应性能相当的滤波器, IIR 比 FIR 的计算量相差悬殊 (见图3.6), 因此 IIR 滤波器主要用于对计算资源有要求、且可以接受非线性相位的场合。

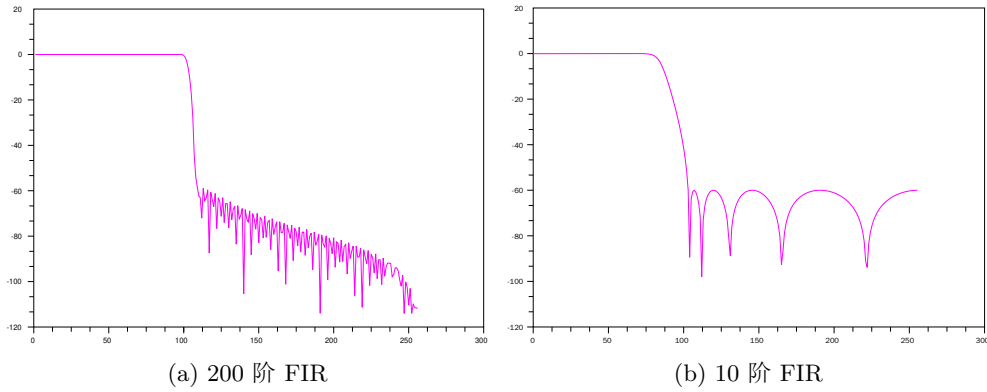


图 3.6: FIR 与 IIR 计算性能对比

3.2.1 零极点配置法

根据一般性 IIR 传递函数

$$H(z) = k \frac{(1 - z_1 z^{-1})(1 - z_2 z^{-1}) \cdots (1 - z_M z^{-1})}{(1 - p_1 z^{-1})(1 - p_2 z^{-1}) \cdots (1 - p_N z^{-1})} \quad (3.10)$$

可以看出, 当需要让系统阻止某些频率时, 可以在该频率处设置零点 z_m , 如果要求增大某些频率的响应, 可在该频率处设置极点 p_n 。零点位置是在 z 平面的单位圆上、角频率 ω_m 处, 即:

$$z_m = \exp(j\omega_m) \quad (3.11)$$

为保证系统的稳定性, 极点必需位于 z 平面的单位圆以内。极点越接近单位圆, 对应角频率 ω_n 处的幅度响应越强:

$$z_n = r \exp(j\omega_n) \quad (3.12)$$

为了避免出现虚数的滤波器系数, 零点或极点必需是实数、或者共轭出现。

根据以上基础, 若要设计一个中心频率 0.2π 的带通滤波器, 可以按下面的步骤:

1. 在 $\omega = 0$ 和 $\omega = \pi$ 处设置两个零点: $z_1 = 1, z_2 = -1$;
2. 在 $\omega = 0.2\pi$ 处设置一个极点: $p_1 = 0.95 \exp(j * 0.2\pi)$;
3. 由于极点是虚数, 需要补充一个共轭极点 \tilde{p}_1 。

图3.7是零极点位置和滤波器幅频响应。

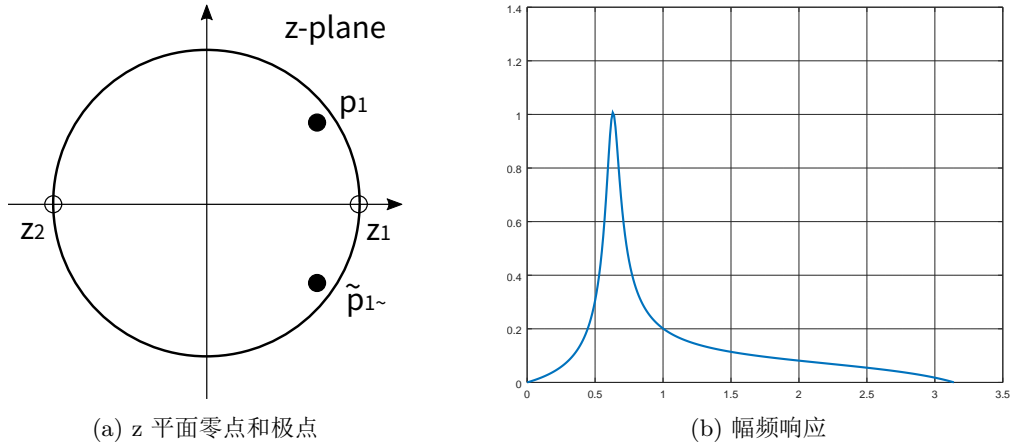


图 3.7: 零极点配置法设计滤波器

为了改善过渡带, 根据带宽的要求, 还应在中心频率附近补充零点和极点。

这种零极点配置法的优点是直接, 不存在模拟-数字域转换, 不足之处是设计手段不成熟, 比较依赖经验。

3.2.2 由模拟滤波器变换

由于我们对模拟滤波器研究得比较透彻, 因此, 从模拟滤波器着手设计数字滤波器是目前的主要手段。

巴特沃斯滤波器

巴特沃斯 (Butterworth) 低通滤波器的一般模型为

$$H(s) = \frac{\Omega_c^N}{(s - s_1)(s - s_2) \cdots (s - s_N)} \quad (3.13)$$

其中 s_k 是 $2N$ 极点, 它们均匀分布在 s 平面左半边以 Ω_c 为半径的圆周上, 即

$$s_k = \Omega_c e^{j\pi \frac{2k+N-1}{2N}}$$

式中 k 取 $1, 2, \dots, N$ 。当 N 为偶数时, s_k 必需共轭出现; 如果 N 为奇数, 则有一个是实数 $-\Omega_c$ 。

(3.13) 式写成

$$H(s) = \begin{cases} \frac{\Omega_c^N}{\prod_{k=1}^{N/2} (s^2 + 2\Omega_c s \sin(\frac{2k-1}{2N}\pi) + \Omega_c^2)}, & N = \text{even} \\ \frac{\Omega_c^N}{(s + \Omega_c) \prod_{k=1}^{(N-1)/2} (s^2 + 2\Omega_c s \sin(\frac{2k-1}{2N}\pi) + \Omega_c^2)}, & N = \text{odd} \end{cases} \quad (3.14)$$

将模拟系统转换到数字域, 目前主要有冲激响应不变法、阶跃响应不变法和双线性变换法。

冲激响应不变法

冲激响应不变法是利用连续时间滤波器来设计离散时间无限冲激响应滤波器的一种方法。这种方法中, 对连续时间系统的冲激响应进行采样以产生离散时间系统的冲激响应。

以 T 为采样周期对连续时间系统的冲激响应 $h_c(t)$ 采样, 得到离散时间系统的冲激响应 $h(n) = Th_c(nT)$, 该系统的频率响应为

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} H_c(j\frac{\omega}{T} + j\frac{2\pi}{T}k)$$

如果连续时间系统的频带被限制在小于采样的奈奎斯特频率的范围内, 则离散时间系统的频率响应则与连续系统的频响一致:

$$H(e^{j\omega}) = H_c(j\frac{\omega}{T}), \text{ for } |\omega| < \pi$$

一个有 N 个极点 $s_k, k = 1, 2, \dots, N$ 的连续系统, 可部分分式展开为:

$$H_c(s) = \sum_{k=1}^N \frac{A_k}{s - s_k} \quad (3.15)$$

通过拉普拉斯变换, 得到因果系统的冲激响应

$$h(t) = \sum_{k=1}^N A_k \exp(s_k t), t \geq 0 \quad (3.16)$$

对应离散系统的冲激响应为

$$h(n) = T \sum_{k=1}^N A_k \exp(s_k nT) u(n) \quad (3.17)$$

将上式 z 变换, 可得其离散时间系统函数

$$H(z) = T \sum_{k=1}^N \frac{A_k}{1 - e^{s_k T} z^{-1}} \quad (3.18)$$

以上可知, 采用冲激响应不变法将模拟系统转换到数字系统, 先将模拟系统展开为部分分式之和, 得到系统的极点 s_k , 映射到离散时间系统的极点 $e^{s_k T}$ 。

双线性变换法

用双线性变换将模拟滤波器变换到数字域, 需要如下的代换

$$s = \frac{1 - z^{-1}}{1 + z^{-1}} \quad (3.19)$$

并将数字域频率进行预畸 (prewarp)

$$\Omega_c = \tan \frac{\pi f_c}{F_s}$$

即: 模拟滤波器指标中的频率参数并不真是给定的参数, 而是通过给定数字域的频率参数按上式进行预畸得到。

采用冲激响应不变法, 模拟系统特性到数字系统的特性保持不变, 包括相位特性。但由于采样定理的限制, 会产生混叠, 因此不适用于高通或带阻型滤波器, 即使对于低通或带通, 也要求高频衰减足够大。双线性变换法则没有这些限制。在不考虑相位因素时, 双线性变换法更具有普适性。以下主要采用双线性变换法。

为简化书写, 对 (3.14) 式中的系数作如下代换

$$p_k = 2 \sin\left(\frac{2k-1}{2N}\pi\right)$$

则每一个二阶部分表示为

$$H(z) = \frac{\Omega_c^2(1 + 2z^{-1} + z^{-2})}{(1 + p_k\Omega_c + \Omega_c^2) + (2\Omega_c^2 - 2)z^{-1} + (1 - p_k\Omega_c + \Omega_c^2)z^{-2}} \quad (3.20)$$

一阶部分

$$H(z) = \frac{\Omega_c + \Omega_c z^{-1}}{(1 + \Omega_c) + (\Omega_c - 1)z^{-1}} \quad (3.21)$$

Chebyshev-I 型低通滤波器

Butterworth 滤波器在通带内是平坦的, $s = 0$ 处的前 N 阶导数为零, 它又叫最大平坦型滤波器。Butterworth 低通滤波器的止带衰减速度只和阶数有关。每增加一阶, 每倍频程衰减增加 6dB。然而, 人为在通带或止带增加一些纹波, 可换取更优的过渡带性能。这一思想导致切比雪夫 (Chebyshev) 型滤波器的诞生。Chebyshev 滤波器的设计思想是, 牺牲通带或止带的一些特性, 以换取过渡带的性能。Chebyshev-I 型允许在通带内的频响有一定的起伏 (纹波), Chebyshev-II 型将纹波设计在止带 (即, 对于低通滤波器来说, 随着频率的增加, 衰减不会持续增加)。II 型 Chebyshev 滤波器较少使用。

N 阶 Chebyshev-I 型低通滤波器幅度响应的平方:

$$|H(j\Omega)|^2 = \frac{1}{1 + \varepsilon^2 C_N^2\left(\frac{\Omega}{\Omega_c}\right)} \quad (3.22)$$

式中 ε 是纹波系数 ($0 < \varepsilon < 1$), ε 越大, 纹波越大。 $C_N(x)$ 是 N 阶 Chebyshev 多项式, 定义为

$$C_N(x) = \begin{cases} \cos(N \cos^{-1} x), & x \leq 1 \\ \cosh(N \cosh^{-1} x), & x > 1 \end{cases} \quad (3.23)$$

表 3.2: Chebyshev 多项式

N	$C_N(x)$
1	x
2	$2x^2 - 1$
3	$4x^3 - 3x$
4	$8x^4 - 8x^2 + 1$
5	$16x^5 - 20x^3 + 5x$
6	$32x^6 - 48x^4 + 18x^2 - 1$

表3.2是前 6 阶切比雪夫多项式。

Chebyshev-I 型低通滤波器的极点的实部和虚部满足椭圆方程。给定通带内纹波系数 δdB , 首先计算

$$\varepsilon = \sqrt{10^{\delta/10} - 1}$$

定义

$$\phi = \frac{1}{N} \sinh^{-1} \frac{1}{\varepsilon}$$

左半平面的极点表示式为

$$s_k = -\Omega_c \sinh \phi \sin\left(\frac{2k-1}{2N}\pi\right) + j\Omega_c \cosh \phi \cos\left(\frac{2k-1}{2N}\pi\right) \quad (3.24)$$

考虑到首项系数的特点, 实际模拟滤波器写成

$$H(s) = \frac{\Omega_c^N}{N \varepsilon 2^{N-1} \prod_{k=1}^N (s - s_k)} \quad (3.25)$$

Chebyshev-I 型低通滤波器止带衰减不仅和阶数有关, 还和通带纹波有关。从给定衰减率计算参数比较复杂, 建议直接设定阶数和纹波。

如将二阶滤波器写成

$$H(s) = \frac{\Omega_c^2}{s^2 + p_k \Omega_c s + q_k \Omega_c^2} \quad (3.26)$$

(p_k, q_k 可由 s_k 的实部和虚部计算得到)

则

$$H(z) = \frac{\Omega_c^2(1 + 2z^{-1} + z^{-2})}{(1 + p_k \Omega_c + q_k \Omega_c^2) + (2q_k \Omega_c^2 - 2)z^{-1} + (1 - p_k \Omega_c + q_k \Omega_c^2)z^{-2}} \quad (3.27)$$

如果允许滤波器通带和止带都有纹波, 可以获得更陡的过渡特性。由于这种滤波器需要用到椭圆函数, 因此被称为椭圆滤波器。图3.8是相同阶数的四种类型滤波器幅频特性对比。这些滤波器在工程实现时, Butterworth 对参数的敏感性最低, 椭圆滤波器对参数的敏感性最高。高敏感参数意味着, 当参数稍有变化时, 滤波器特性会发生较大的变化, 在有限字长条件下, 应慎重选择。

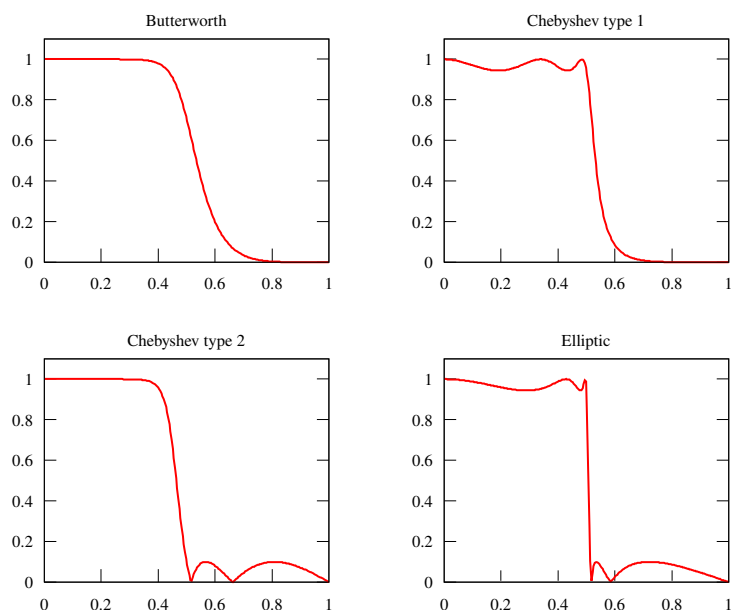


图 3.8: 不同类型滤波器幅频特性

Bessel 低通滤波器

贝塞尔滤波器是具有最大平坦的群延迟 (线性相位响应) 的线性滤波器, 它以德国数学家 Friedrich Bessel 命名。贝塞尔低通滤波器从通带到止带的过渡相对比较缓慢, 但在几乎整个通频带都具有恒定的群延迟 (线性相位), 因而在通频带上保持了被过滤的信号波形。图3.9是 4 阶 Bessel 低通滤波器的幅频响应和群延迟。

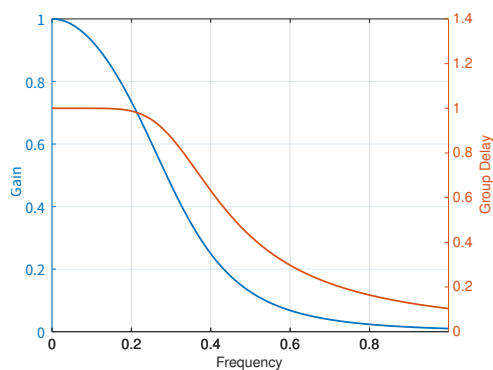


图 3.9: 4 阶 Bessel 低通滤波器的频响特性

按 Ω_c 归一化 Bessel 低通滤波器的一般公式为

$$H(s) = \frac{q_0}{q_0 + q_1 s + q_2 s^2 + \dots q_n s^n} \quad (3.28)$$

其分母系数的通项是 $q_k = \frac{(2n-k)!}{2^{n-k} k! (n-k)!}$, $q_0 = q_1$

以下是前 6 阶 Bessel 低通的归一化 ($\Omega_c = 1$) 原型:

$$\begin{aligned}
 H(s) &= \frac{1}{1+s} \\
 H(s) &= \frac{3}{3+3s+s^2} \\
 H(s) &= \frac{15}{15+15s+6s^2+s^3} \\
 H(s) &= \frac{105}{105+105s+45s^2+10s^3+s^4} \\
 H(s) &= \frac{945}{945+945s+420s^2+105s^3+15s^4+s^5} \\
 H(s) &= \frac{10395}{10395+10395s+4725s^2+1260s^3+210s^4+21s^5+s^6}
 \end{aligned}$$

表3.3是 1~10 阶 Bessel 模拟滤波器的极点。由这些极点可以生成二阶 (一阶) 数字形式的级联公式。

根据以上极点参数表, 模拟 Bessel 滤波器 (非归一化) 可写成

$$H(s) = \frac{q_0 \Omega_c^n}{\prod_{k=1}^n (s - \Omega_c s_i)} \quad (3.29)$$

如果滤波器是偶数阶, 则每二阶可以写成

$$H(s) = \frac{\Omega_c^2}{s^2 + p_k \Omega_c s + q_k \Omega_c^2} \quad (3.30)$$

最后再乘以一个系数 q_0 . 如果是奇数阶, 再乘以一项实极点一阶因子

$$\frac{\Omega_c}{s - s_n \Omega_c} \quad (3.31)$$

用双线性变换公式 $s = \frac{1-z^{-1}}{1+z^{-1}}$ 将上面的二阶模拟低通变换到数字域 (模拟频率先经预畸 $\Omega_c = \tan \frac{\pi f_c}{F_s}$):

$$H(z) = \frac{\Omega_c^2(1+2z^{-1}+z^{-2})}{(1+p_k \Omega_c + q_k \Omega_c^2) + (2q_k \Omega_c^2 - 2)z^{-1} + (1-p_k \Omega_c + q_k \Omega_c^2)z^{-2}} \quad (3.32)$$

3.2.3 滤波器类型变换

对于高通、带通等其他类型的滤波器, 可以在低通滤波器的基础上进行简单的变换得到。表3.4是其他类型滤波器的代换公式。表中 Ω_p 是截止频率, Ω_{pc} 是中心频率 (几何中心), B_p 是带宽。

高通滤波器的形式

$$H(z) = \frac{1-2z^{-1}+z^{-2}}{(\Omega_c^2 + p_k \Omega_c + q_k) + (2\Omega_c^2 - 2q_k)z^{-1} + (\Omega_c^2 - p_k \Omega_c + q_k)z^{-2}} \quad (3.33)$$

(和低通形式相比, 分子 z^{-1} 系数变号, p_k 和 q_k 的计算公式不变。)

表 3.3: Bessel 滤波器极点参数表

阶数	极点 s_i	P_k	Q_k
1	-1.0000000000		
2	-1.5000000000±0.8660254038i	3.0000000000	3.0000000000
3	-1.8389073227±1.7543809598i -2.3221853546	3.6778146454	6.4594326935
4	-2.1037893972±2.6574180419i -2.8962106028±0.8672341289i	4.2075787944 5.7924212056	11.4878004771 9.1401308900
5	-2.3246743032±3.5710229203i -3.3519563992±1.7426614162i -3.6467385953	4.6493486064 6.7039127984	18.1563153132 14.2724805136
6	-2.5159322478±4.4926729537i -3.7357083563±2.6262723114i -4.2483593959±0.8675096732i	5.0318644956 7.4714167126 8.4967187918	26.5140253444 20.8528231769 18.8011305897
7	-2.6856768789±5.4206941307i -4.0701391636±3.5171740477i -4.7582905282±1.7392860611i -4.9717868585	5.3713537578 8.1402783272 9.5165810564	36.5967851564 28.9365460928 25.6664447530
8	-2.8389839489±6.3539112986i -4.3682892172±4.4144425005i -5.2048407906±2.6161751526i -5.5878860433±0.8676144454i	5.6779678978 8.7365784344 10.4096815812 11.1757720866	48.4320186525 38.5692532753 33.9347400844 31.9772252587
9	-2.9792607982±7.2914636883i -4.6384398872±5.3172716754i -5.6044218195±3.4981569179i -6.1293679043±1.7378483835i -6.2970191817	5.9585215964 9.2768797744 11.2088436390 12.2587358086	62.0414376213 49.7885026571 43.6466457530 40.5892679102
10	-3.1089162336±8.2326994591i -4.8862195669±6.2249854825i -5.9675283286±4.3849471889i -6.6152909655±2.6115679208i -6.9220449054±0.8676651955i	6.2178324672 9.7724391338 11.9350566572 13.2305819310 13.8440898108	77.4427005313 62.6255859132 54.8391562020 50.5823615631 48.6675485637

Butterworth 高通滤波器

从模拟低通滤波器到高通的变换, 只需用 $\frac{\Omega_c^2}{s}$ 替换式 (3.14) 中的 s . 由此导出高通数字滤波器公式:

表 3.4: 归一化低通到其他类型的变换

目标类型	代换式
低通	$\bar{s} = \frac{s}{\Omega_p}$
高通	$\bar{s} = \frac{\Omega_p}{s}$
带通	$\bar{s} = \frac{s^2 + \Omega_{pc}^2}{B_p s}$
全通	在 s 平面右半边增加与极点对应的零点

二阶部分

$$H(z) = \frac{1 - 2z^{-1} + z^{-2}}{(1 + p_k \Omega_c + \Omega_c^2) + (2\Omega_c^2 - 2)z^{-1} + (1 - p_k \Omega_c + \Omega_c^2)z^{-2}} \quad (3.34)$$

一阶部分

$$H(z) = \frac{1 - z^{-1}}{(\Omega_c + 1) + (\Omega_c - 1)z^{-1}} \quad (3.35)$$

Bessel 高通滤波器

模拟域用 $\frac{\Omega_c^2}{s}$ 替换 s , 可以得到高通形式. 再经双线性变换得到数字高通:

$$H(z) = \frac{1 - 2z^{-1} + z^{-2}}{(\Omega_c^2 + p_k \Omega_c + q_k) + (2\Omega_c^2 - 2q_k)z^{-1} + (\Omega_c^2 - p_k \Omega_c + q_k)z^{-2}} \quad (3.36)$$

Bessel 滤波器形式和 Chebyshev-I 型完全一样. 只是在给定纹波系数的条件下, Chebyshev-I 型滤波器的参数可以解析表示, 而 Bessel 通用多项式的根无解析表达式, 只能通过数值方法求解。

二阶带通滤波器

二阶带通由一阶归一化低通滤波器变换得到。

$$H(s) = \frac{1}{s + 1} \quad (3.37)$$

将式中的 s 换成 $s + \frac{\Omega_0}{s}$, Ω_0 是带通滤波器的几何中心,

$$\Omega_0 = \sqrt{\Omega_1 \cdot \Omega_3}$$

Ω_1, Ω_3 分别是带通滤波器的上下截止频率 (3dB 点).

带通滤波数字形式为

$$H(z) = \frac{g(1 - z^{-2})}{(D + 1) - D \cdot E z^{-1} + (D - 1)z^{-2}} \quad (3.38)$$

其中,

$$\begin{aligned} D &= \operatorname{ctg}(B_w T/2) \\ E &= \frac{2 \cos(\omega_0 T)}{\cos(B_w T/2)} \end{aligned}$$

参数说明如下:

g 增益 (线性)

ω_0 中心角频率

B_w 带宽 (角频率), 和 Q 值的关系为 $Q = \omega_0/B_w$

T 采样周期

3.2.4 峰值滤波/陷波

这种滤波器类似于带通或带阻¹ 模拟二阶峰值滤波器/陷波器 (Peak/Notch) 的一般模型

$$H(s) = \frac{s^2 + 2g\alpha\Omega_c s + \Omega_c^2}{s^2 + 2\alpha\Omega_c s + \Omega_c^2} \quad (3.39)$$

显然, $H(j0) = H(j\infty) = 1$, $H(j\Omega_c) = g$, 且 $s = \Omega_c$ 处的一阶导数为零, 中心角频率 Ω_c 处的增益为 g 。当 $g > 1$ 时为 Peak, 而当 $g < 1$ 时则为 Notch。通带 (或阻带) 临界频率以 \sqrt{g} 为参照, 不再采用 3dB 带宽指标。这类滤波器可以以较小的运算量做出比较窄的峰值滤波或陷波。图3.10是不同参数下滤波器的幅频特性。

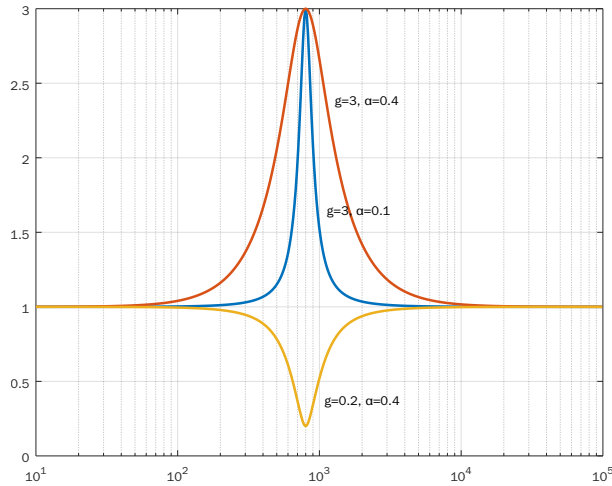


图 3.10: Peak/Notch 型滤波器幅频特性

和带宽有关的参数是 α 。经过双线性变换, 离散数字域的形式写成:

$$H(z) = \frac{(1 + ga) - 2dz^{-1} + (1 - ga)z^{-2}}{(1 + a) - 2dz^{-1} + (1 - a)z^{-2}} \quad (3.40)$$

¹Notch 的特性与带阻类似, 但 Peak 并不是带通, 它在中心频率处有正的增益, 但在其他频率没有衰减。

其中

$$\begin{aligned} a &= \frac{1}{\sqrt{g}} \sinh\left(\frac{\ln 2}{2} \cdot \frac{\omega_0}{\sin \omega_0} \cdot b_w\right) \sin(\omega_0 T) \\ d &= \cos(\omega_0 T) \end{aligned} \quad (3.41)$$

b_w 是倍频程带宽, 对于上、下限频率 ω_+ 和 ω_- , 满足

$$\begin{aligned} \omega_+ &= 2^{b_w} \omega_- \\ Q &= \frac{1}{2^{b/2} - 2^{-b/2}} \end{aligned}$$

Peak/Notch 型滤波器常用于多通道均衡器。

3.2.5 IIR 滤波器结构选型

从 IIR 滤波器一般形式

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^N b_k z^{-k}}{1 - \sum_{k=1}^M a_k z^{-k}} \\ &= \frac{b_0 + b_1 z^{-1} + \dots + b_N z^{-N}}{1 - a_1 z^{-1} - \dots - a_M z^{-M}} \end{aligned} \quad (3.42)$$

得到差分方程:

$$y(n) = \sum_{k=0}^N b_k x(n-k) + \sum_{k=1}^M a_k y(n-k)$$

图3.11是直接 I 型, 对应上面的差分方程。

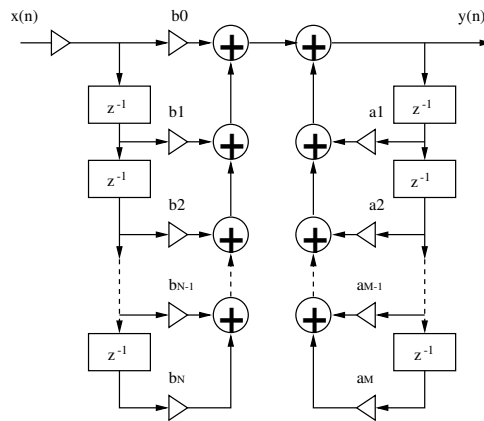


图 3.11: IIR 直接 I 型结构

将直接 I 型两个级联子系统交换顺序, 就成为直接 II 型。实用时会将延迟单元合并。通常 $M \geq N$, 直接 II 型可以使用较少的延迟单元。如果是硬件实现, 意味着较少的寄存器开销。直接型的缺点是, 系数 a_k 、 b_k 对滤波器性能控制作用不明显, 且极点对系数比较敏感。

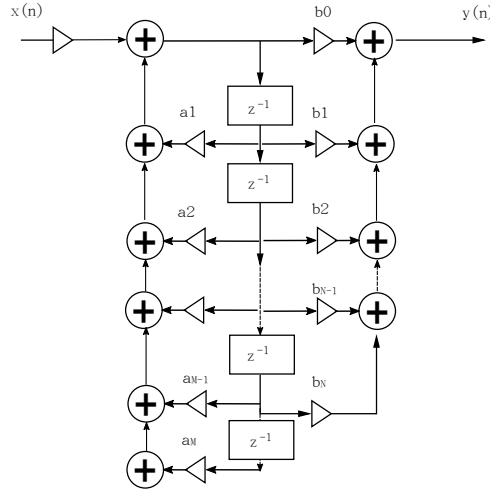


图 3.12: IIR 直接 II 型结构

将式 (3.42) 按零极点因式分解, 考虑到实系数多项式可以分解成实系数一次项与二次项的乘积, 得到:

$$H(z) = A \prod_k \frac{1 + \beta_{1k}z^{-1} + \beta_{2k}z^{-2}}{1 - \alpha_{1k}z^{-1} - \alpha_{2k}z^{-2}} \prod_k \frac{1 + \beta_k z^{-1}}{1 - \alpha_k z^{-1}} \quad (3.43)$$

它由若干一阶、二阶子系统级联而成 (图3.13)。级联的特点是可以独立调整各级的零极点。

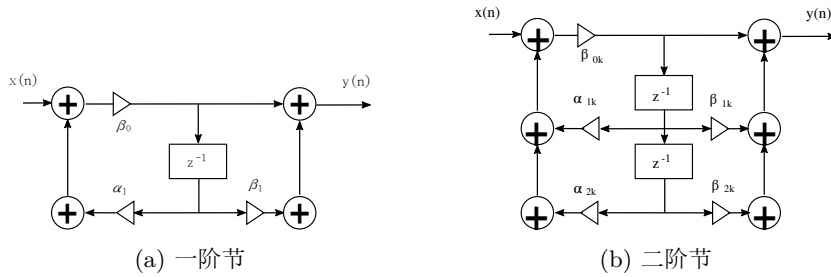


图 3.13: IIR 级联型

在级联结构中, 分子、分母可以形成不同的组合。在有限字长计算中, 不同的搭配会导致误差结果的不同, 这是一个值得研究的课题。此外, 系统总增益 A 需要分散到各级, 以避免某一级增益过大产生溢出, 或者某一级过小导致信噪比太低。

式 (3.42) 还可以展开成部分分式形式。当多项式分子的次数不低于分母时, $H(z)$ 可写成如下形式:

$$H(z) = \sum_{k=1}^{N_1} \frac{A_k}{1 - c_k z^{-1}} + \sum_{k=1}^{N_2} \frac{\gamma_{0k} + \gamma_{1k} z^{-1}}{1 - \alpha_{1k} z^{-1} - \alpha_{2k} z^{-2}} \quad (3.44)$$

如果分子阶数高于分母 ($N > M$), 上式应再并联一项 FIR 结构。

并联结构可以单独调整每一节的极点, 且各节之间的误差影响较小。

3.3 梳状滤波器

一些系统需要去除特定单频信号 (如 50Hz 工频干扰), 梳状滤波器可以较好地满足要求。当干扰信号周期是采样周期的整倍数 L 时, 一个整周期累加可以将干扰信号抹平, 即:

$$y(n) = \sum_{k=0}^{L-1} x(n-k)$$

该系统函数可以写成

$$\begin{aligned} H(z) &= 1 + z^{-1} + z^{-2} + \cdots + z^{-L+1} \\ &= \frac{1}{1 - z^{-L}} \end{aligned}$$

该算法简化成

$$y(n) = x(n) + y(n-L)$$

这个滤波器严格滤除周期为 L 的信号 (及其谐波), 对于邻近频率的信号没有任何影响。由于频响特性过于尖锐, 一个比较常用的形式是:

$$H(z) = b \frac{1 - z^{-L}}{1 - \rho z^{-L}} \quad (3.45)$$

式中 $b = (1 + \rho)/2, 0 < \rho < 1$ 。 ρ 越大, 频率特性越尖锐。通过选择适当的 ρ , 可使滤波器有一定的带宽。

上面的滤波器对特定频率及其谐波频率造成衰减。梳状滤波器还有一种互补形式, 它对特定频率及其谐波有增强作用。其传递函数是

$$H(z) = b \frac{1 + z^{-L}}{1 - \rho z^{-L}} \quad (3.46)$$

图3.14是两种梳状滤波器的频响特性。

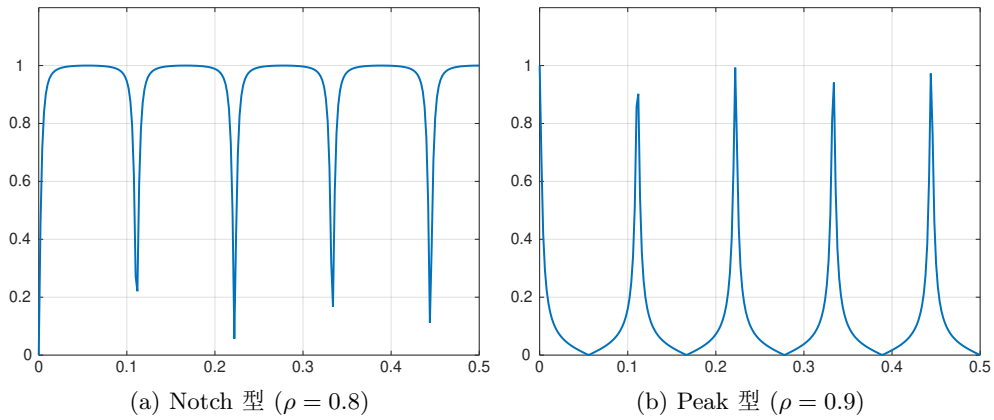


图 3.14: 梳状滤波器幅频响应

4

自适应滤波器

4.1 维纳滤波

4.1.1 问题的提出

许多实际问题中, 数字信号处理的目的是抑制干扰、突出有用信号。例如消除“回声”型干扰信号。这种干扰信号基本上和原信号一样, 只是在幅度和相位上有所不同 (见图4.1)。因此, 频

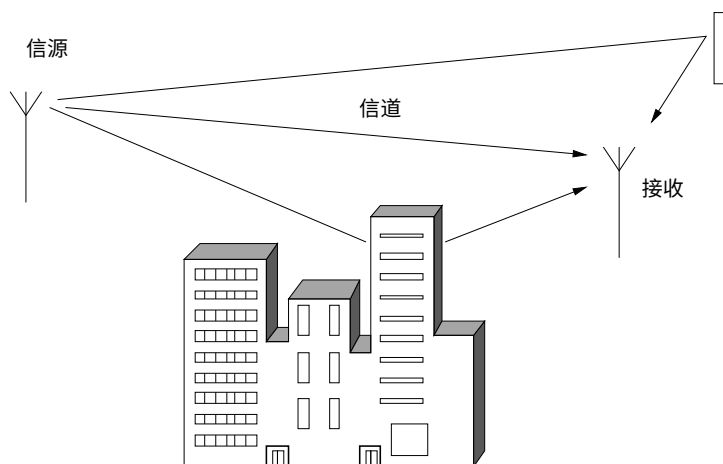


图 4.1: 回波反射干扰

率滤波器 (如低通、高通、带通、点阻滤波器等) 无法实现回波消除, 因为频率滤波器适用于那些有用信号与干扰信号在频谱上具有一定分离特性的场合。而“回声”型干扰与原始信号在频谱上是一致的, 不具备可分离特性。

为解决“回声”或“重影”消除问题, 人们提出了“最小平方滤波的概念”。最小平方滤波的设计和实现涉及三种信号 (见图4.2):

- 输入信号 $x(n)$
- 期望 (理想) 输出信号 $d(n)$
- 实际输出信号 $y(n)$

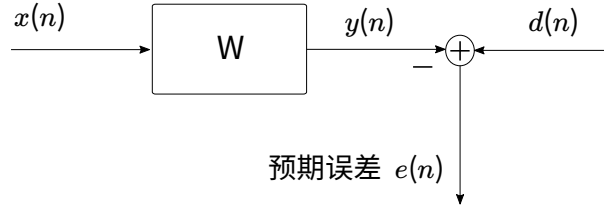


图 4.2: 最小平方滤波

最小平方准则要求

$$E = \sum_n [d(n) - y(n)]^2$$

最小, 而最小平方滤波器称为维纳滤波器, 维纳滤波器输出

$$y(n) = \sum_{k=0}^N w(k)x(n-k)$$

以此准则找出的滤波因子进行的滤波为最小平方滤波。

上式对滤波因子 $w(n)$ 求导, 得到一组方程

$$\sum_{k=0}^N r_{xx}(k-j)w(k) = r_{xd}(j), \quad j = 0, 1, \dots, N \quad (4.1)$$

这里

$$r_{xx}(m) = \sum_n x(n-m)x(n) \quad (4.2)$$

$$r_{xd}(m) = \sum_n x(n-m)d(n) \quad (4.3)$$

分别为 $x(n)$ 的自相关序列和 $x(n)$ 与 $d(n)$ 的互相关序列。

以最小平方滤波理论为基础的线性误差滤波对于消除混响型干扰有很好的效果。对于另外一种类型的干扰——随机噪声干扰的消除问题 (例如, 从含噪语音中抑制噪声, 突出语音信号), 由于语音的频谱 (如辅音) 与噪声非常相似, 不可能用频率滤波器将噪声与语音分离。此时, 最小平方滤波也难以奏效。抑制这类噪声问题的理论基础是随机信号处理理论。最小平方滤波理论中的“相关序列”、“不相关”、“可预测”等概念的精确定义或准确说明, 只能建立在随机信号处理的理论基础上。维纳滤波理论实际上是最小平方滤波理论的深入, 它将最小平方滤波方法从确定信号处理提高到了平稳随机信号处理阶段。

到目前为止, 我们有了以下干扰信号的处理基础:

1. 对有限频带干扰, 可以采用频率滤波器;
2. 处理回声重影这类的干扰, 最小平方滤波不失为一种有效的方法;¹
3. 随机噪声干扰的消除, 是建立在随机信号处理理论基础上的, 到目前为止尚无普遍适用的手段。

¹消除“回声”、“混响”型干扰还可以采用同态信号处理的方法。

4.1.2 Wiener-Hopf 方程

维纳²滤波的一般模型与最小平方滤波的一般模型基本一致, 只是将后者的所有有关信号换成为数学期望为零、相互之间为平稳相关的平稳随机信号

维纳滤波器是一个有限冲激响应的滤波器, 希望输出信号和实际输出信号二者之间越接近越好, 即要求误差

$$e(n) = d(n) - y(n)$$

均方误差意义上达到最小。在给定滤波器阶数 $M + 1$ 的前提下, 寻找一组最佳滤波因子

$$w_o(0), w_o(1), \dots, w_o(M)$$

使得它们在所有可能的滤波因子中能使均方误差 (Mean-Square Error) $E[e^2(n)]$ 取最小值。我们称使 $E[e^2(n)]$ 为最小的寻求最佳滤波因子的准则为最小均方误差准则。

采用 FIR 型滤波器设计, 令

$$\mathbf{w} = [w(0), w(1), \dots, w(M)]^T$$

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M)]^T$$

式中 $[]^T$ 表示矩阵转置。该滤波器的输出为

$$y(n) = \mathbf{w}^T \mathbf{x}(n) = \mathbf{x}^T(n) \mathbf{w}$$

均方误差

$$\begin{aligned} E[e^2(n)] &= E[(d(n) - y(n))^2] \\ &= E[(d(n) - \mathbf{w}^T \mathbf{x}(n))(d(n) - \mathbf{x}^T(n) \mathbf{w})] \end{aligned} \quad (4.4)$$

设 $J(\mathbf{w}) = E[e^2(n)]$, 这里 $J(\mathbf{w})$ 称为价值函数。寻找 \mathbf{w}_o , 使

$$J(\mathbf{w}_o) = \min J(\mathbf{w})$$

最终可推导出求解最佳滤波因子 \mathbf{w}_o 的 Wiener-Hopf³ 方程的解

$$\mathbf{w}_o = \mathbf{R}^{-1} \mathbf{p} \quad (4.5)$$

其中 \mathbf{R} 是 $x(n)$ 的自相关矩阵, \mathbf{p} 是序列 $x(n)$ 和 $d(n)$ 的互相关序列。与 LPC 方法类似, Wiener-Hopf 方程也采用 Levison-Durbin 算法求解。

² Norbert Wiener (1894.11.26 – 1964.3.18), 美国数学家, MIT 教授, 控制论的先驱。

³ Eberhard Frederick Ferdinand Hopf (1902.4.4 – 1983.7.24), 德国数学家。

4.2 自适应滤波

4.2.1 最速下降法

在解决维纳滤波问题时, 按照最小均方误差准则导出

$$J(\mathbf{w}) = \sigma_d^2 - 2\mathbf{p}^T \mathbf{w} + \mathbf{w}^T R \mathbf{w} \quad (4.6)$$

价值函数 $J(\mathbf{w})$ 是多维曲面的二次函数。当滤波器为二阶时, $J(\mathbf{w})$ 是开口向上的抛物面。求解维纳滤波因子 \mathbf{w}_o 就是求 $J(\mathbf{w})$ 的最小值问题。形象地说就是找到误差性能曲面的“碗底”坐标。

从多元微积分知道, 用 $J(\mathbf{w})$ 关于 \mathbf{w} 的各个分量的偏导数构成的矢量 $\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}$ 称为 $J(\mathbf{w})$ 的梯度 $\nabla J(\mathbf{w})$ 。在 $\mathbf{w} = \mathbf{w}_0$ 处的梯度是 $J(\mathbf{w})$ 的值在 \mathbf{w}_0 附近增加最快的方向, 在误差性能曲面上 $(\mathbf{w}_0, J(\mathbf{w}_0))$ 最陡的方向。反方向 $-\nabla J(\mathbf{w})$ 是 $J(\mathbf{w})$ 值的最速下降方向。以此方法求解价值函数最小值的方法被称为最速下降法。

最速下降法从 \mathbf{w} 的初始值出发

$$\mathbf{w}_0 = [w_0(0), w_0(1), \dots, w_0(M)]^T$$

沿负梯度方向移动一小步, 得到新的 \mathbf{w} 值,

$$\mathbf{w}_1 = [w_1(0), w_1(1), \dots, w_1(M)]^T$$

迭代公式

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{1}{2}\mu \nabla J(\mathbf{w}_n) \quad (4.7)$$

反复进行这样的迭代, 可以证明, 当迭代步长 μ 足够小时, 我们就可以用这种迭代方法求出 \mathbf{w}_o 。

4.2.2 最小均方自适应滤波算法

维纳滤波问题中, 要寻求 Wiener-Hopf 方程的最佳滤波因子。在解这个方程时, 需要知道方程的系数矩阵 R 和向量 \mathbf{p} ($x(n)$ 的自相关和 $x(n)$ 与 $d(n)$ 的互相关), 即要求掌握随机信号 $x(n)$ 和 $d(n)$ 的先验信息。只有在这些信息准确的基础上才能保证获得准确的维纳滤波因子 \mathbf{w}_o , 从而获得良好的滤波效果。

而在实际应用中, 这是一个苛刻的要求。通常的做法是用随机序列的有限长时间平均

$$\begin{aligned} \hat{r}_{xx}(l) &= \sum_{n=0}^N x(n)x(n+l) \\ \hat{r}_{xd}(l) &= \sum_{n=0}^N x(n)d(n+l) \end{aligned}$$

来取代总体平均 $r_{xx}(l)$ 和 $r_{xd}(l)$, 然后计算出滤波因子进行滤波。

这样做的问题在于:

1. 用一段时间平均代替总体平均, 如要获得更高的精度, 就需要有更大的 N , 势必增加时延, 这对实时信号处理是很不利的;

2. 随机序列的平稳只是相对的, $r_{xx}(l)$ 和 $r_{xd}(l)$ 往往只是在一段时间内可以认为不随时间而变。而实际上, 随着时间的推移、外界环境等因素的影响终究导致 $r_{xx}(l)$ 和 $r_{xd}(l)$ 的变化。这种变化立即会使滤波的效果变差, 因为用以前的数据估计出的 $r_{xx}(l)$ 和 $r_{xd}(l)$ 不适用于后来的信号。

解决这两个问题的有效途径是采用自适应滤波方法。在进行滤波运算时, 不断地调整滤波因子, 以使它们适应输入信号的变化, 从而取得良好的滤波效果。

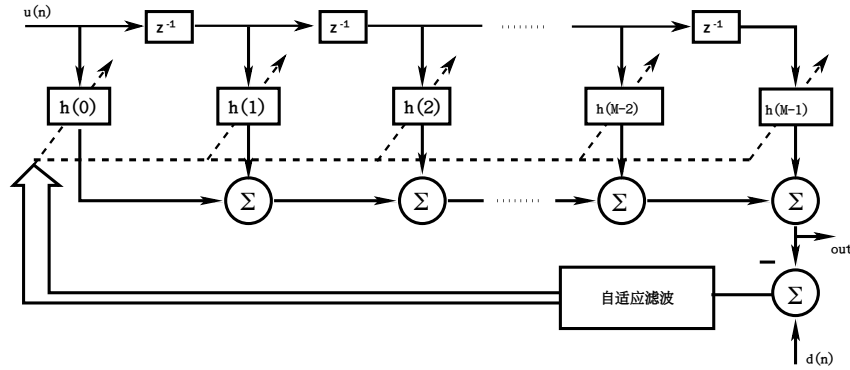


图 4.3: FIR 自适应滤波模型

自适应滤波包含两个基本过程:

1. 滤波过程, 包括
 - 计算横向滤波器的输出;
 - 与期望响应比较, 产生估计误差。
2. 自适应过程, 根据估计误差自动调整滤波器系数。

自适应滤波的过程, 从一组预先指定的滤波因子出发, 经过逐次迭代的调整 (这种调整依据实际输出 $y(n)$ 与期望输出 $d(n)$ 之间的差异, 用某个自适应控制算法, 如最速下降法、最小均方法等), 在不知道随机信号序列 $x(n)$ 和 $d(n)$ 的确切相关特性的情况下, 平稳环境中滤波因子能够收敛于最佳的 Wiener 滤波因子, 即使在某些非平稳环境下, 自适应滤波过程也有自动跟踪的能力。当随机信号的相关特性缓慢变化时, 滤波因子也能跟着变化, “适应” 环境, 得到较好的滤波效果。

在求解 Wiener-Hopf 方程中, 可能会用到牛顿迭代法、最速下降法。这些算法都需要预先知道 $x(n)$ 的相关特性。更为实用的是最小均方自适应滤波算法。LMS (Least Mean Square) 自适应滤波方法被认为是把维纳滤波理论潜力在实用中充分发挥出来的非常成功的途径, 目前在通信、雷达、声纳、语音处理等领域中取得了很好的应用效果。

在推导最速下降法中, 需要短时平均值 $E\{e^2(n)\}$ 进行梯度估计。只需要用 $e^2(n)$ 本身代替短时平均值, 便得到 LMS 算法的迭代公式

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e(n) \mathbf{x} \quad (4.8)$$

上式中看出, 在实际系统中实现 LMS 算法不需要平方、平均或者微分运算, 算法的优美正在于它的简易性和效率。梯度矢量的每一分量由单个数据样本得到。不用平均, 梯度分量肯定包含了噪声成份。但是在自适应过程中, 从某种意义上讲, 过程的进行起着一个小通滤波的作用, 随着迭代的进行, 噪声会逐渐衰减。

4.2.3 自适应滤波的应用

自适应噪声抵消器

图4.4是一个双输入自适应噪声抵消器的算法框图。信号 s 沿信道传到接收该信号的传感器。除收到信号外, 传感器还收到一个不相关的噪声 x' 。组合信号和噪声 $s + x'$ 构成抵消器的主通道信号。第二个传感器接收与目标信号不相关、但以某种未知的方式与噪声 x' 相关的噪声 x , 作为参考通道输入。将该输出从主通道输出中减去, 产生系统输出 $s + x' - y$ 。

如能知道噪声的传输特性, 则一般而言, 就可以设计能够将 x 变成 $y = x'$ 的固定滤波器。然而, 由于传输通道的特性是未知的, 或者是变化的, 很难采用固定参数的滤波器。

现采用一个自适应滤波器来处理参考输入信号。采用恰当的算法, 该系统可以在变化的环境条件下工作, 并可不断地调节自身, 使输出误差最小。

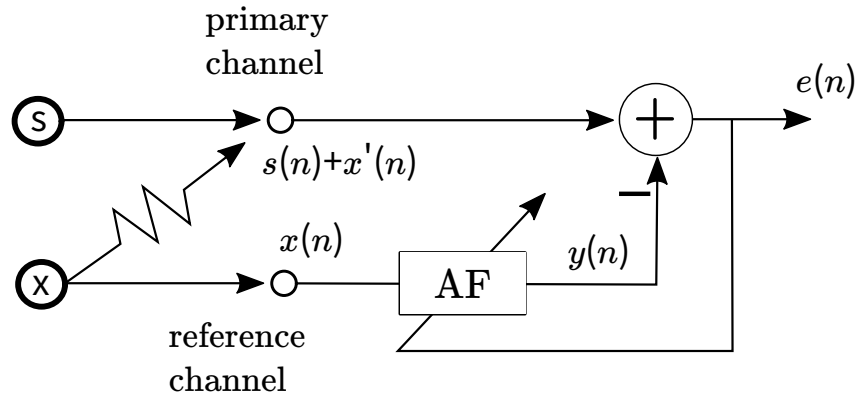


图 4.4: 噪声抵消

图4.4中 y 是自适应滤波器 AF 的输出。当 AF 采用 FIR 结构时,

$$y(n) = \sum_k h(k)x(n-k)$$

系统的输出

$$e = s + x' - y$$

上式两边平方, 得

$$e^2 = s^2 + (x' - y)^2 + 2s(x' - y)$$

两边取期望值。假定信号 s 与干扰噪声 x 都是平稳的, x' 与 x 相关, 且与 s 不相关, 得

$$E[e^2] = E[s^2] + E[(x' - y)^2]$$

当调节滤波器使 $E[e^2]$ 最小时, 信号功率 $E[s^2]$ 不受影响, 而使 $E[(x' - y)^2]$ 达到最小。

另一方面, 如果参考输入 x 与主通道中的噪声 x' 完全不相关, 这种情况下

$$\begin{aligned} E[e^2] &= E[(s + x')^2] - 2E[y(s + x')] + E[y^2] \\ &= E[(s + x')^2] + E[y^2] \end{aligned}$$

要使输出功率最小, 就要求 $E[y^2]$ 最小, 这可以令滤波器所有系数 $h(k)$ 为零来实现, 从而关闭自适应滤波器。即: 自适应滤波器不会使输出的效果变坏。

自调谐滤波器

自适应噪声抵消器的另一种应用是构成自调谐滤波器。图4.5是自调谐滤波器的工作原理。我们感兴趣的信号是周期信号, 或是具有较强相关性的信号。而干扰信号通常是宽带噪声。通过延迟, A 处的宽带干扰信号与 B 处的原信号相关性将大大降低, 如果周期信号仍能保持相关性, 则可以利用自适应噪声抵消器的原理将干扰信号去除。

如果周期信号的周期已知, 增强周期信号也可以用梳状滤波器, 如果周期未知, 或存在多个复合周期, 梳状滤波会变得比较复杂。此时自调谐滤波器可以很好地解决这一问题。自适应滤波器稳定后, 滤波器呈现的频谱特性与梳状滤波器一致。

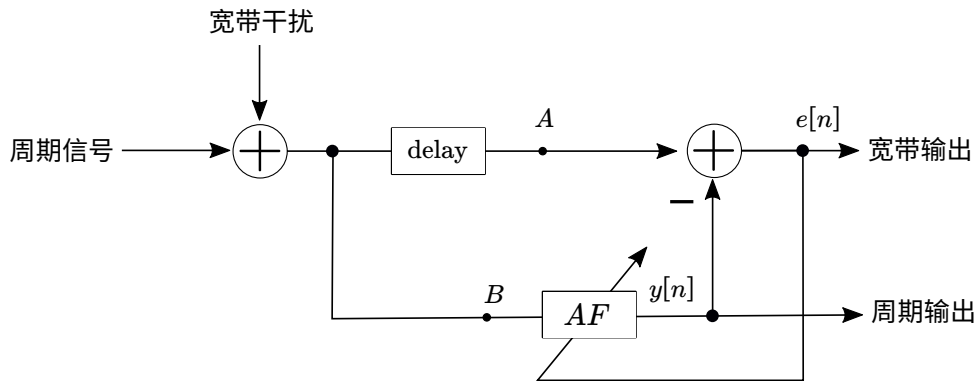


图 4.5: 自调谐滤波器

一定场合下, 图4.5模型也可以解决相反的问题, 即抑制周期信号 (像是 50Hz 工频干扰), 提取其他的有用信号。这时候, 有效输出可取自 $e(n)$ 。

回声抵消

回声是语音通信中常见的现象, 普通对话环境也会有室内回声。当存在扩音设备时, 回声还将导致电声系统啸叫。图4.6是电话通信回声产生的模型。远端说话人 (Speaker B) 将话音通过通信线路传到近端 (Speaker A) 的扬声器输出, 如果没有回声处理系统, 说话人 A 的声音混合 B 的声音会叠加后, 经过 A 端的传声器传到远端, 说话人 B 会同时听到 A 和自己的声音。由于说话人 A 所在的声环境是不确定的, 不能用固定滤波的方法处理回声, 此时自适应滤波就可以发挥作用了。

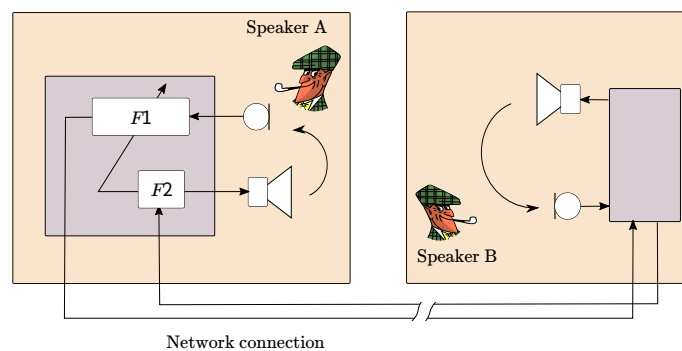


图 4.6: 电话通信中的回声产生模型

图4.7是自适应电话回声抵消系统。该算法在电话回声抵消、啸叫处理上均得到了较好的应用。

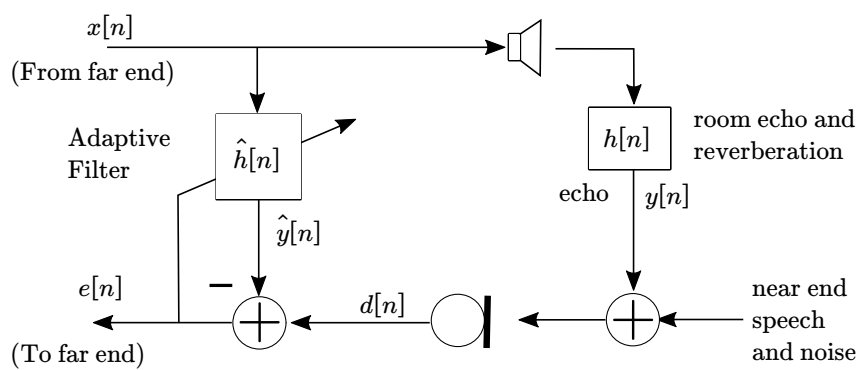


图 4.7: 自适应回声抵消系统

5

音频工程应用



5.1 工程计算中的定点小数处理

数字信号处理涉及计算密集型算法。而大量的应用场合又要求低成本、低功耗, 因此开发出了很多快速算法, 以便在性能不太高的处理器中实现算法功能。

在工程应用领域, 浮点数可以实现很高的精度, 而定点数的精度却非常受限制。从经济的角度考虑, 定点处理器则更具优势。但由于浮点格式相当复杂, 在定点处理器上运算时会产生很大的计算开销。矛盾就是这样产生的。解决的方法之一是依赖技术的进步, 降低浮点处理器的价格成本, 目前看来确实有这样的趋势。另一方面, 通过算法改进, 将浮点运算变为定点运算, 在一段时间里也是一种手段。

考虑如下的 FIR 算法:

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (5.1)$$

若设计滤波器时得到的系数 $h(n)$ 是浮点数格式, 取值范围为 $(-8, +8)$; 数字信号 $x(n)$ 通过 16bit 量化, 取值范围为 $[-32768, +32767]$ 的定点数; 计算结果 $y(n)$ 可直接送往 16bit DAC 进行模拟输出, 它也应在 $[-32768, +32767]$ 范围内。为此设计如下计算方法:

1. 将 $h(n)$ 乘以 2^{12} , 用 16 位存储单元保留整数部分, 小数部分作为误差舍弃。记住, 整数部分已经将原本的 $h(n)$ 放大了 2^{12} 倍, 可以认为最高 4 位是 $h(n)$ 的整数部分 (首位为符号位, 实际上整数部分只有 3 位), 低 12 位是 $h(n)$ 的小数部分。这种表示格式被称为 Q12.3。本质上, 这种数据格式是定点数, 小数点的位置只是人为规定的, 不需要做任何计算处理。
2. 类似的, 对于 $x(n)$, 简单认定小数点在最高位之后 (最高位是符号), 此谓 Q15 格式。
3. 由于 $h(n)$ 与 $x(n)$ 都已是定点数, 可直接在定点处理器中通过一条定点指令完成一次 $h(k) \cdot x(n-k)$ 的乘法计算。16 位的定点处理器中, $16\text{bit} \times 16\text{bit}$ 的结果保存在 32bit 单元中。根据乘法运算规则, 结果的高 5 位是整数部分 (两个数的符号位各占 1 位), 低 27 位是小数部分。
4. 由于 $h(n)$ 比实际数字增大了 2^{12} 倍, 如果要将结果还原到 Q15 格式, 还应将 $y(n)$ 除以 2^{12} , 也就是向右移 12 位。

考虑到加法运算可能产生溢出, 式 (5.1) 的中间过程可以保留 32 位 (Q27.4 格式), 最后再转换到 Q15。

5.2 单频信号谱分析

确定一个信号的频率是数字信号处理的一个重要应用。使用有限长序列分析单频信号的频率会面临频率分辨率精度的问题。图5.1是不同长度分析 100Hz 采样 20Hz 正弦信号的结果。采样率 F_s 的 N 点 DFT, 频率分辨率为 F_s/N 。当信号频率恰好在分辨率点上时才能得到正确的结果 (图5.1(b)), 而加大 N 有助于提高分辨率。

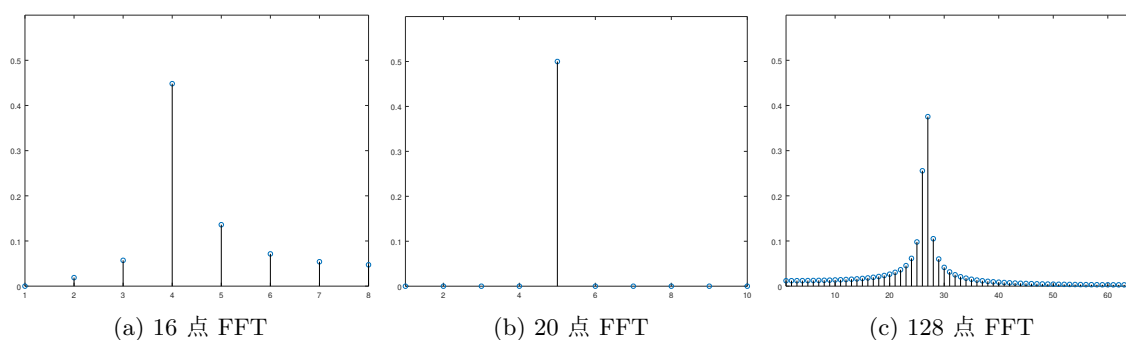


图 5.1: 不同长度频率分析结果

信号处理理论告诉我们, 有限长信号分析存在频率泄露。改善泄露的一个方法是使用合适的窗函数截断信号。窗函数的主瓣应该窄, 而旁瓣幅度应该小。

存在多个信号叠加时, 问题会更加复杂。图5.2是接近的两个频率不同的分析方法: 增加分析点数和选择窗函数。32 点窗长度不足以分辨两个频率, 而 64 点 FFT 的分辨率达不到。只有两个条件都满足时才可以将两个接近的频率分开。

$$x(n) = 0.85 \sin(\omega_1 n) + \sin(\omega_2 n), \omega_1 = 0.22\pi, \omega_2 = 0.26\pi \quad (5.2)$$

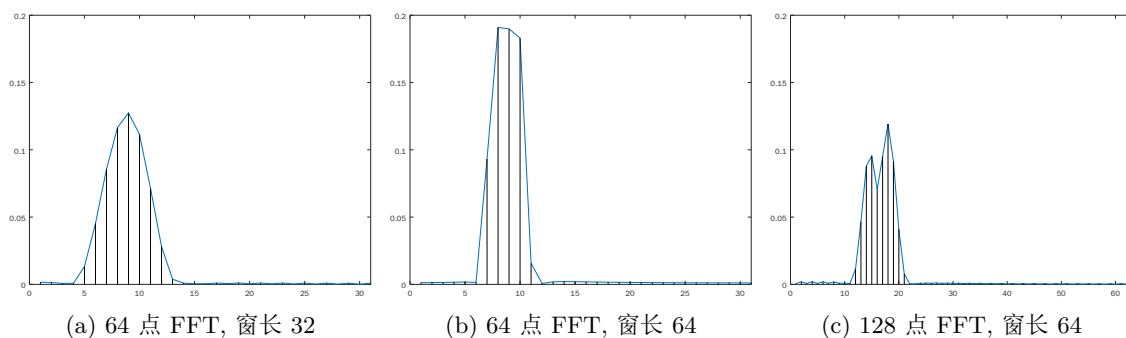


图 5.2: 不同长度、窗长的频率分析结果

结论: 基于 DFT 的谱分析, 需要考虑三个因素: 一、选择适当的窗函数, 尽可能窄的主瓣和尽可能小的旁瓣; 二、增加窗长度可以减小主瓣宽度; 三、增加分析时间可以增加频率分辨率, 频率分辨率是时间的倒数。

5.3 双音多频检测

按键式电话拨号比早期的拨盘式脉冲电话更快。按键电话采用双音多频技术 (Dual Tone Multiple Frequency — DTMF)。DTMF 还用于语音控制、电话银行、自动取款机等设备。双音多频信号由两个单音合成, 这两个单音各来自一组预设频率信号。一对音频代表唯一的数字或符号。因此, 经过双音多频解码即可识别它们对应的符号。

	1209Hz	1336Hz	1477Hz	1639Hz
697Hz	1	2	3	A
770Hz	4	5	6	B
852Hz	7	8	9	C
941Hz	*	0	#	D

图 5.3: DTMF 拨号盘

由于待检测的频率点有限 (只有 8 个), 并且存在频谱泄露问题, 利用 FFT 算法检测双音多频没有优势。DTMF 解码采用 Goertzel¹ 算法。Goertzel 算法使用二阶滤波器计算输入序列 $x(n)$ 的第 k 个 DFT 系数:

$$s_k(n) = x(n) + 2 \cos\left(\frac{2\pi k}{N}\right) s_k(n-1) - s_k(n-2) \quad (5.3)$$

$$\begin{aligned} y_k(n) &= s_k(n) - W_N^k s_k(n-1) \\ &= s_k(n) - e^{-j2\pi k/N} s_k(n-1) \end{aligned} \quad (5.4)$$

N 个采样点处理后便得到第 k 个 DFT 系数, 即

$$X(k) = y_k(n)|_{n=N} \quad (5.5)$$

对应功率值

$$X(N)X^*(N) = s_k^2(N) - 2 \cos\left(\frac{2\pi k}{N}\right) s_k(N)s_k(N-1) + s_k^2(N-1) \quad (5.6)$$

对于 8kHz 采样率, N 取 205, 表5.1是对应不同频率的 k 的取值。

图5.4是拨号“0696675356”的频谱分析和 Goertzel 算法结果。10 个键的拨号时间共约 1 秒, 采样频率 8kHz。相比于 DFT, Goertzel 算法在 DTMF 检测上有很多优势: 它只需要计算有限的频率点, 节省了计算量; 不像 FFT 算法, 对 N 的长度不要求是 2 的整幂次; 每个滤波器只有两级延迟线, 节省存储空间。

¹ Gerald Howard Goertzel (1919.8.18 — 2002.7.17), 美国理论物理学家, 1958 年提出 Goertzel 算法。

表 5.1: DTMF 取值

音调 (Hz)	样点位置	整数 k
697	17.861	18
770	19.731	20
852	21.833	22
941	24.113	24
1209	30.981	31
1336	34.235	34
1477	37.848	38
1639	41.846	42

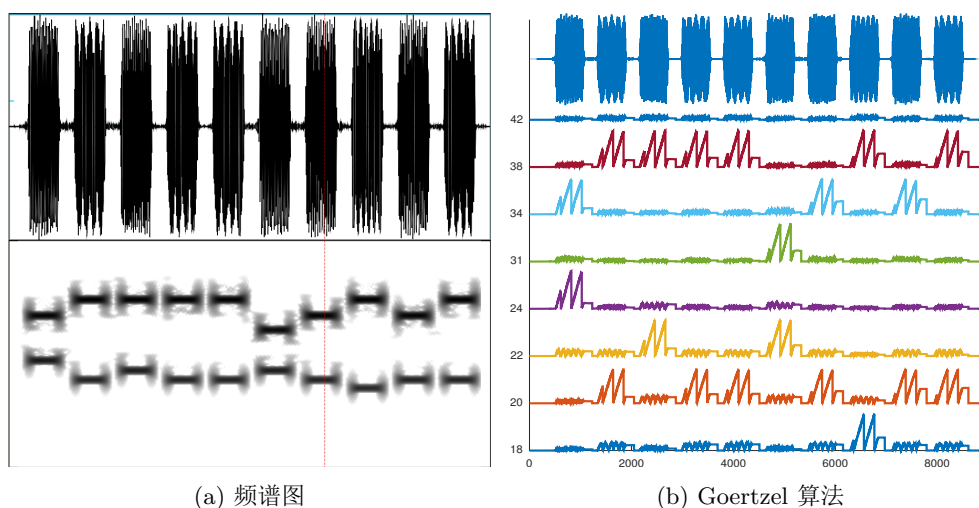


图 5.4: Goertzel 算法

5.4 回声与混响

产生回声最简单的办法是数字延迟。延迟通常也是产生更复杂的结果的基础。图5.5系统可以产生一次回声。延迟时间可设置在 10ms 到 30ms 之间。考虑到声反射面总要吸收一些能量, 这里的 $a < 1$ 。它的传递函数是

$$H(z) = 1 + az^{-D} \quad (5.7)$$

多反射回声可按一次回声结构增加多个延迟节实现, 也可以考虑采用反馈型结构 (图5.6)。

在一个较大的听音室里, 混响声场更为复杂 (图5.7)。到达听音者的声音包含直达声、早期反射声和混响三部分。早期反射集中在前 80ms, 大多数由壁面一次反射构成, 可通过 FIR 结构模拟; 六个梳状滤波器模仿混响声场, 图5.8延时单位基于 44.1kHz 采样率建立。

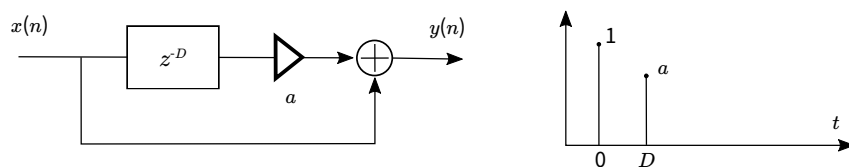


图 5.5: 一次回声及其冲激响应

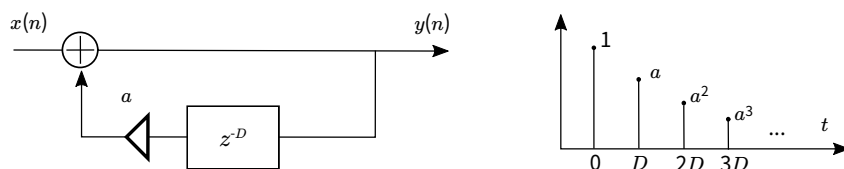


图 5.6: 反馈型多反射回声效果

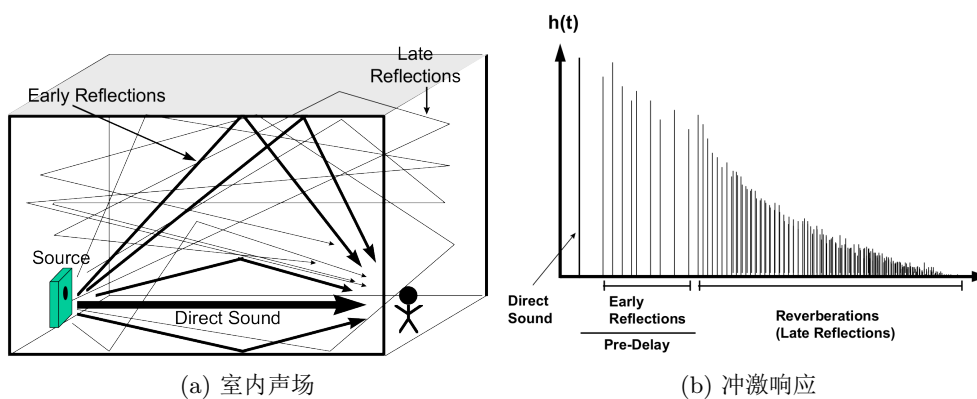


图 5.7: 大型室内声场环境

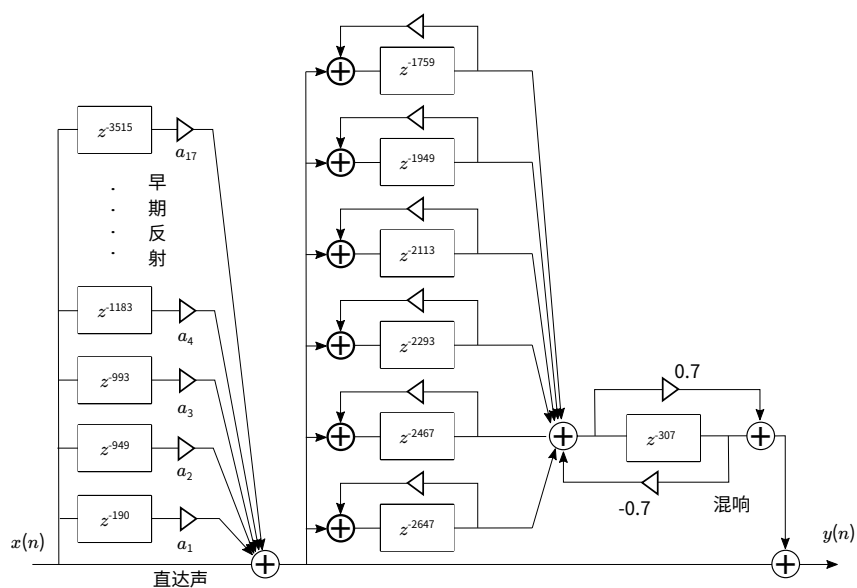


图 5.8: 室内声场仿真

表 5.2: FIR 模拟早期反射 (波士顿交响音乐厅数据,Delay in ms)

Tap	Delay	Gain	Tap	Delay	Gain
0	0	1			
1	4.3	0.841	10	58.7	0.193
2	21.5	0.504	11	59.5	0.217
3	22.5	0.490	12	61.2	0.181
4	26.8	0.379	13	70.7	0.180
5	27.0	0.380	14	70.8	0.181
6	29.8	0.346	15	72.6	0.176
7	45.8	0.289	16	74.1	0.142
8	48.5	0.272	17	75.3	0.167
9	57.2	0.192	18	79.7	0.134

表 5.3: 梳状滤波器特性 (Delay in ms)

	Delay	g
1	40	0.32
2	44	0.31
3	48	0.30
4	52	0.30
5	56	0.28
6	60	0.26
Allpass	7	0.7

5.5 变声

变声器可以改变乐器或声音的音调，在音乐中产生合唱 (Chorus) 或颤音 (Vibrator) 效果。一些匿名采访中，变声器用来让匿名者的声音无法辨认。这几种音响效果的处理方法大同小异，主要差别在延迟线的方式。变声器采用锯齿波延迟。

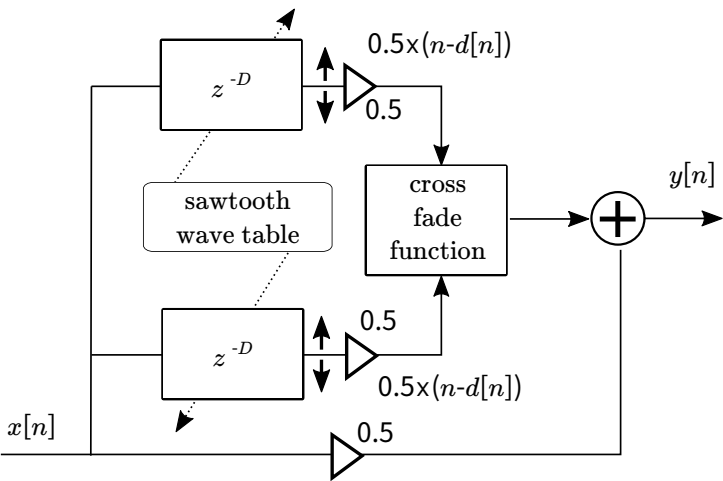


图 5.9: 带有渐变的变声器

由于延迟线在缓冲区中移动的速度不定，单延迟线算法的变声在延迟线指针变化时会产生“click”声。图5.9是带有过渡渐变 (cross fade) 的延迟方法。Cross fade 可采用简单的三角窗函数在两个通道中平移实现。

5.6 压限

压限器 (Compressor/Limiter) 是音响设备的常用组件。它的主要目的是扩展音频的动态幅度特性。在一个线性系统中, 当输入信号过大时, 放大器的输出会产生上下限幅而导致严重失真, 甚至可能烧毁设备。压限器会在此时有意改变放大器的放大倍数, 减少这种失真。图5.10表现某种音响设备器输入-输出的关系, 在直线转折点以上的斜率被称为压限比。

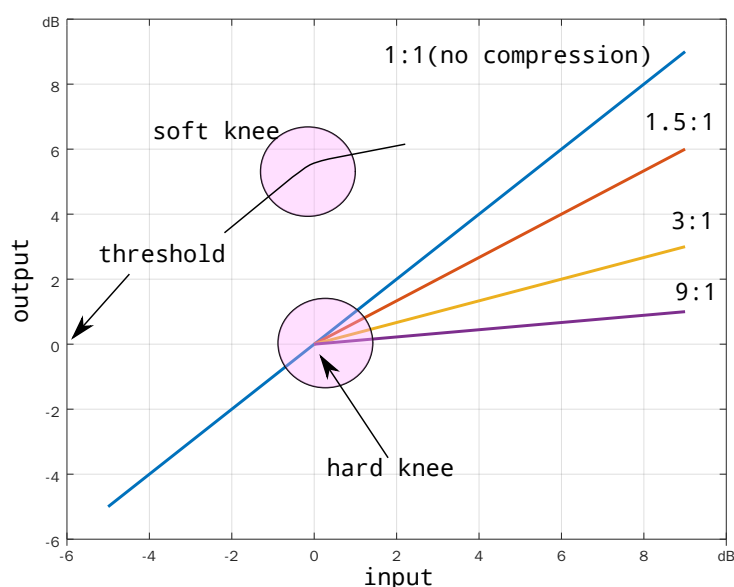


图 5.10: 压限器输入-输出幅度关系

压缩比是压限器的重要参数, 它表示当输入电平超过门限后输出与输入的增量比。这个值越大, 动态幅度消减得越厉害。当压缩比为 $\infty : 1$ 时, 即为 Limiter (一般超过 20:1 即认为 Limiter)。

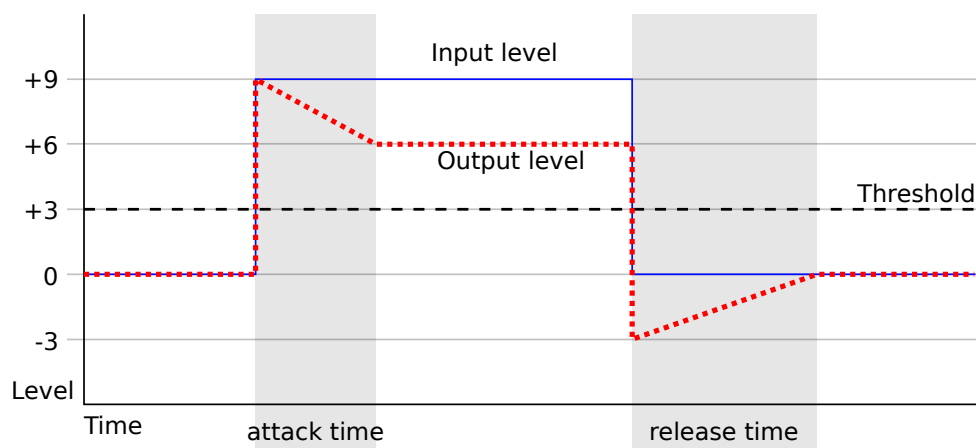


图 5.11: 压限器动态输出

压限器并不是以信号超过门限的瞬时时间工作, 而是会有一个滞后; 同样, 信号回落到门限以下, 压限器的压缩比也会滞后一段时间才回到 1:1。前者称为启动时间 (attack time), 通常在 50ms 左右, 后者称为恢复时间 (release time), 通常被设置在 500ms 左右。专业调音师会根据音

源特性 (如音乐的节拍、乐器的衰减特性) 改变这些参数。

5.7 阵列信号处理

5.7.1 多路输出系统

音响设备中可以看到大小不同的扬声器。由于目前技术水平所限, 单个扬声器很难做到覆盖人耳的频率感知范围 (20Hz–20kHz)。普遍采取的做法是, 用不同的扬声器还原不同的频率。通常尺寸大的扬声器低频响应好, 而小尺寸的扬声器适合高频响应。电声器材将 20Hz–20kHz 频率范围的信号通过滤波器分隔成不同的频段, 以适用后级扬声器的频响。

5.7.2 阵列扬声器

在一些扩声场景中, 需要对声场进行有目的的控制。传统的机械条件下, 可以用人工制作的反射物形成声聚焦或者声散射。前者用于形成定向声场, 后者的目的是产生均匀声场 (图5.12)。在数字化时代, 更多的是采用数字阵列的方法实现声场控制。这样的系统被称为参数阵列扬声器 (Parametric Array Loudspeaker, PAL)。

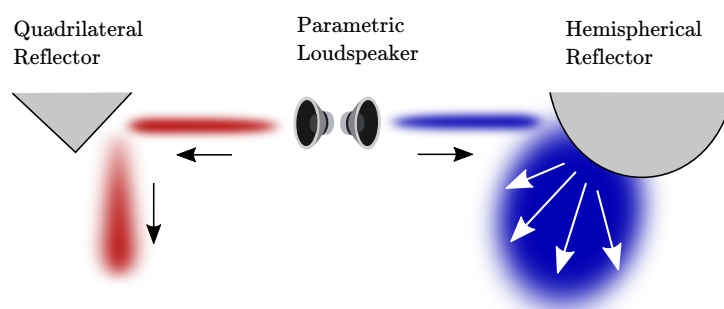


图 5.12: 通过聚焦反射和散射改变声场

阵列扬声器是由一组单元箱组成, 这些单元箱按一定规定排列。单元箱的辐射特性包括辐射声功率、频率特性、水平指向性、相位等参数。

阵列扬声器常面临以下几种应用场景:

1. 提高远距离声场强度。当把单元扬声器视为点声源时, 声强满足平方反比定律。在大型室外场馆扩声设备需要考虑远场强度时常常通过构造线阵列扬声器实现。
2. 室内 (音乐厅、会议厅等) 扩声需要考虑声场的均匀性。
3. 形成特定的指向性。

以可调指向性阵列 (steerable PAL) 为例, 图 5.13 展示了延迟–累加波束成型 (Delay-and-sum beamforming) 结构。²

² Chuang Shi; Yoshinobu Kajikawa and Woon-seng Gan, *An overview of directivity control methods of the parametric array loudspeaker*, APSIPA Transactions on Signal and Information Processing, 3, E20.

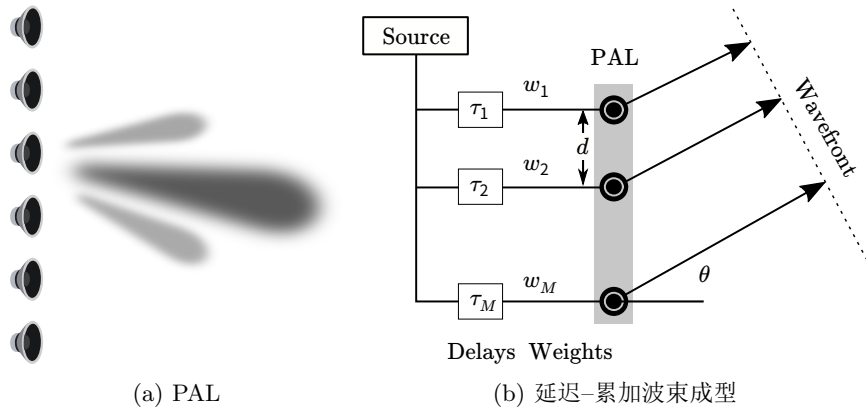


图 5.13: PAL 及其声场分析

系统包含等间距 d 布置的 M 个无指向性扬声器, 构成一个线阵。通常将两个相邻的扬声器设置为相等的延迟 τ , 即第 m 个扬声器的延迟为

$$\tau_m = m \frac{d}{c_0} \sin \theta_0, \text{ for } m = 0, 1, \dots, M-1$$

(式中 c_0 为声速, θ_0 为聚焦方向角)

在聚焦方向角以外, 不同频率指向性特性是不一样的。假定声压按线性传播:

$$p(x, t) = \sum_m \exp \left[-\alpha_m x + j\omega \left(t - \frac{x}{c_0} \right) \right] \quad (5.8)$$

指向性理论值

$$D(\theta) = \left| \sum_m \bar{w}_m \exp \left[j \frac{w_m}{c_0} m d (\sin \theta - \sin \theta_0) \right] \right| \quad (5.9)$$

5.7.3 阵列传声器

阵列传声器希望解决的问题是, 定向接收目标, 抑制来自其他方向 (或者其他位置) 的干扰信号, 提高接收信号的信噪比。阵列传声器接收模型最早是从雷达天线阵列接收装置引入的。它由若干个接收器对信号进行空间“采样”, 形成一个尖锐的指向性, 指向感兴趣的目标。最初的形式也仅仅是延迟-累加 (Delay-and-sum beamforming), 即通过每个接收器的延迟形成接收指向, 最后通过累加, 利用噪声信号的不相关性或者指向性外的信号相位的随机性, 完成对干扰的抑制。

这个模型的主要局限在于对干扰源没采取任何处理措施。为使其能够对干扰环境有所响应, 必须使其有这样的自适应特性: 对干扰源方向接收性能为零 (或是响应尽可能小), 并能自动实时地跟踪干扰特性。据此对该模型进行了改进: 为每个接收器增加了一个可调增益, 并进而发展成一个滤波系统, 通过自适应方法调整权系数, 由此提高了系统输出的信噪比。这种算法成功地用

doi:10.1017/ATSIP.2014.18.

Shi, C.; Gan, W.S. *Grating lobe elimination in steerable parametric loudspeaker*. IEEE Trans. Ultrason., Ferroelectr. Freq. Control, 58 (2011), 437–450.

于窄带信号的抗干扰系统, 如雷达和声纳。基于此模型, 在音频信息处理中派生出两种有代表性的算法, 分别由 Otis Lamont Frost 和 Rainer Zelinski 在 1972 年和 1988 年提出。以下分别称 Frost 模型³和 Zelinski 模型⁴。

Frost 模型

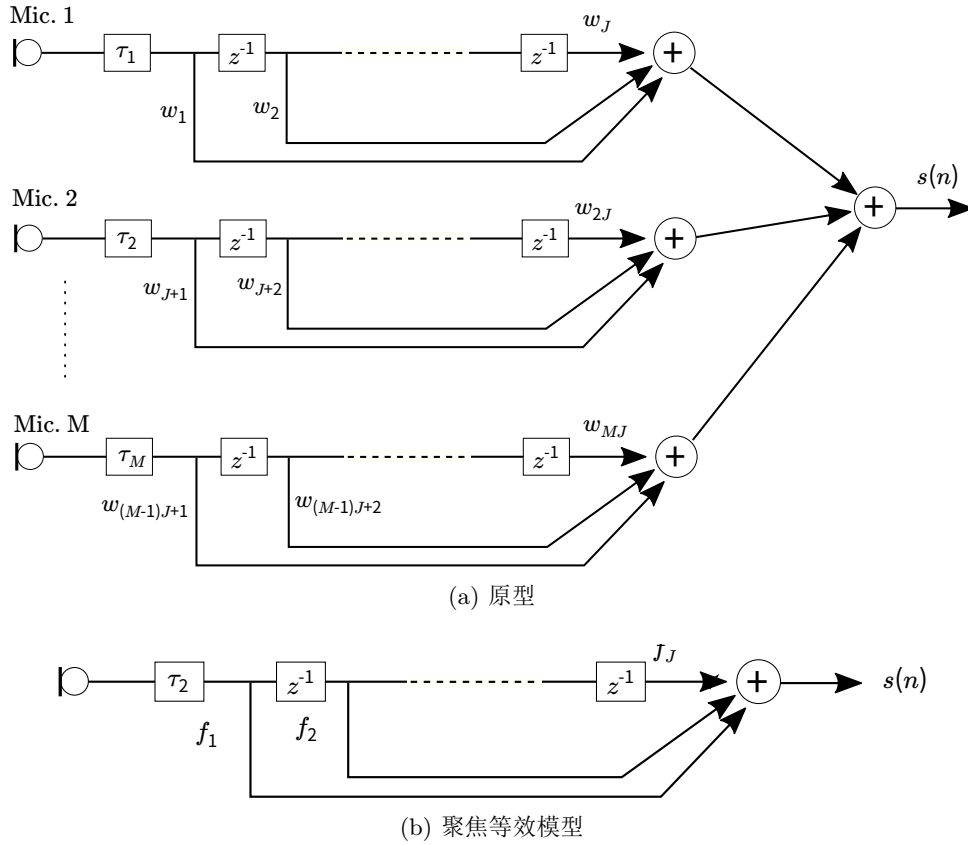


图 5.14: 阵列传声器模型之一—Beamforming

图 5.14(a) 是 Frost 模型示意图。该算法使用 M 个传声器构成接收阵列, 每个传声器后接一个 J 阶 FIR 滤波器, 在一定的约束条件下, 通过极小化输出信号的能量, 确定一组滤波器的系数 w_{mj} , 从而提高信噪比。当要求系统聚焦于信号源方向时, 可以通过调节每个传声器后面的延迟 τ_i 加以实现。5.14(b) 是它相对于信号源方向的等效, 它相当于具有如下频率响应的滤波器:

$$H(z) = \sum_{n=1}^J f_n z^{-n} \quad (5.10)$$

³ Otis Lamont Frost, III, *An Algorithm for Linearly Constrained Adaptive Array Processing*, Proceedings of the IEEE, Vol.60, No.8, August 1972.

⁴ Rainer Zelinski, *A Microphone Array with Adaptive Post-filtering for Noise Reduction in Reverberant Room*, IEEE ICASSP-88, 1988

作为特例, 滤波器系数 f_n 在当 $\tau = \tau_1 = \tau_2 = \cdots = \tau_M$ 时可以写成:

$$f_n = \sum_{m=1}^M w_{mn} \quad (5.11)$$

传声器接收的信号由选定方向的语音信号和来自其他方向的噪声组成。构成的滤波器组应具有这样的特性: 阵列接收指向选定方向, 并尽可能保证其传递函数在这个方向上平坦。

指向性主要靠每路传递函数的延迟实现, 而系统的畸变定义如下:

$$D(\mathbf{w}) = \int_{\Omega} |H(e^{j\omega}) - 1|^2 d\omega \quad (5.12)$$

它应该控制在较小的范围下, 同时又要使其输出的短时能量

$$E(\mathbf{w}) = \frac{1}{N} \sum_{n=0}^{N-1} \{\mathbf{x}^T(n) \cdot \mathbf{w}\}^2 \quad (5.13)$$

达到最小, 从而抑制住来自其他方向的噪声 (形成空间响应的零点)。式中 \mathbf{w} 和 $\mathbf{x}(n)$ 是滤波器系数和输入信号的向量形式:

$$\mathbf{w} = [w_1, w_2, \cdots, w_J, w_{J+1}, \cdots, w_{MJ}]^T \quad (5.14)$$

$$\mathbf{x}(n) = [x_1(n), x_1(n-1), \cdots, x_1(n-J+1), \\ x_2(n), \cdots, x_M(n-J+1)]^T \quad (5.15)$$

为了平衡畸变和短时能量的最小化要求, 引入一个平衡参数 λ , 使 J 最小化:

$$J(\mathbf{w}) = E(\mathbf{w}) + \lambda D(\mathbf{w}) \quad (5.16)$$

可以证明, 系统在空间上可以构成 $M-1$ 个零点, 抑制来自这些方向的干扰信号; 同时在频域上构成 J 个零点, 降低噪声的频谱。较大的 λ 可以保证较好的系统响应, 而较小的 λ 能获得较高的信噪比提高。这种算法在噪声源不多且房间的混响时间短 (墙壁反射系数小) 时比较有效。

Zelinski 模型

Zelinski 模型见图 5.15。硬件结构与 Frost 模型类似, 但算法结构不同。它通过 M 个传声器接收的信号, 估计出信号和噪声的频谱, 由此确定维纳滤波器的系数, 通过维纳滤波实现语音增强的目的。

设每个传声器接收的信号可写为

$$r_i[t] = s[t] + n_i[t], \quad i = 1, \cdots, M \quad (5.17)$$

其中 $s[t]$ 为信号, $n_i[t]$ 为噪声。当每路信号中的噪声互不相关时, 信号的功率谱可由各路信号的互相关估计出:

$$\begin{aligned} R_{ij}[\tau] &= E\{r_i[t] \cdot r_j[t]\} \\ &= R_{ss}[\tau], \quad \text{for } i \neq j \end{aligned} \quad (5.18)$$

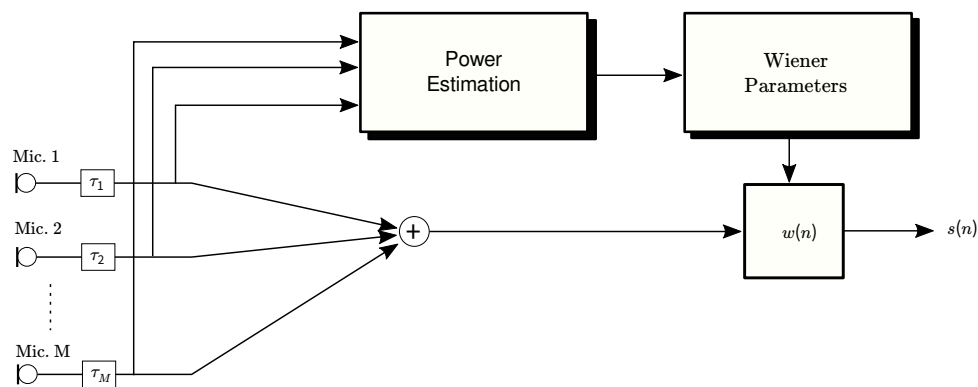


图 5.15: Zelinski 模型

每个传声器接收信号的自相关:

$$\begin{aligned}
 R_{ii}[\tau] &= E\{r_i[t + \tau]r_i[t]\} \\
 &= R_{ss}(\tau) + R_{nn}[\tau]
 \end{aligned} \tag{5.19}$$

由此可建立维纳滤波器

$$W(f) = \frac{\hat{G}_{ij}(f)}{\hat{G}_{ii}(f)} \tag{5.20}$$

式中 $\hat{G}_{ii}(f)$ 为自相关谱, $\hat{G}_{ij}(f)$ 为互相关谱。

Zelinski 算法在噪声源较多, 或混响时间较长的环境下效果比较好, 因为此时的噪声相关性更低。从传声器形式上看, Zelinski 和 Frost 构造是相同的, 但 Zelinski 算法不产生空间指向性。