第3章 语音信号的特征分析

CHAPTER 3

前面讨论了语言学、汉语语音学和信号模型等基础知识。语音信号处理虽然包括语音通信、语音合成、语音识别等，但其前提是对语音信号的分析。只有将语音信号分析表示成其本质特性的参数，才有可能利用这些参数进行高效的语音通信，才能建立用于语音合成的语音库，也才可能建立用于识别的模板或知识库。而且，语音合成的音质好坏、语音识别率的高低，都取决于对语音信号分析的准确性和精度。例如，利用线性预测分析来进行语音合成，其先决条件是要先用线性预测方法分析语音库，如果线性预测分析获得的语音参数较好，则用此参数合成的语音音质就好。又如，利用带通滤波器组法来进行语音识别，其先决条件是要弄清楚语音共振峰的幅值、个数、频率变化范围及其分布情况。因此，应先对语音信号进行特征分析，得到提高语音识别率的有用数据，并据此来设计语音识别系统的硬件和软件。

国内外的经验说明，语音分析的工作必须先于其他的语音信号处理工作。例如，20世纪40年代，贝尔实验室的研究人员就对语音信号分析做了大量的、卓有成效的工作，这些成果推动了语音信号处理的发展。

根据所分析的参数不同，语音信号分析可分为时域、频域、倒谱域等方法。进行语音信号分析时，最先接触到的、最直观的是它的时域波形。语音信号本身就是时域信号，因而时域分析是最早使用且应用范围最广的一种方法。时域分析具有简单直观、清晰易懂、运算量小、物理意义明确等优点，但更为有效的分析多是围绕频域进行的，因为语音中最重要的感知特性反映在其功率谱中，而相位变化只起着很小的作用。

常用的频域分析方法有带通滤波器组方法、傅里叶变换法和线性预测分析法等，其中线性预测方法将在第4章中具体介绍。频谱分析具有如下优点：时域波形较易随外界环境变化，但语音信号的频谱对外界环境变化具有一定的顽健性。另外，语音信号的频谱具有非常明显的声学特性，利用频域分析获得的语音特征具有实际的物理意义，如共振峰参数、基音周期参数等。

倒谱域是将对数功率谱进行反傅里叶变换后得到的，它可以将声道特性和激励特性有效地分开，因此可以更好地揭示语音信号的本质特征。

按照语音学的观点，可将语音信号分析分为模型分析法和非模型分析法两种。模型分析法是指依据语音信号产生的数学模型，来分析和提取表征这些模型的特征参数；共振峰模型分析及线性预测分析即属于这种方法。凡不进行模型化分析的其他方法都属于非模型

分析法，包括上面提到的时域分析法、频域分析法及同态分析法等。

贯穿于语音信号分析全过程的是“短时分析技术”。根据对语音信号的研究，其特性是随时间而变化的，所以它是一个非稳态过程。但从另一方面看，虽然语音信号具有时变特性，但不同的语音是由人的口腔肌肉运动构成声道的某种形状而产生的响应，而这种肌肉运动频率相对于语音频率来说是缓慢的，因而在一个短时间范围内，其特性基本保持不变，即相对稳定，所以可以将其看作是一个准稳态过程。基于这样的考虑，对语音信号的分析和处理必须建立在“短时”的基础上，即进行“短时分析”。将语音信号分为一段一段来分析，其中每一段称为一“帧”(frame)。由于语音信号通常在 10～30ms 之内是保持相对平稳的，因而帧长一般取 10～30ms。

本章首先介绍语音信号的数字化处理，接着介绍语音信号的时域处理技术及频域和倒谱域的相应处理。此外，还将介绍常见的倒谱特征、基音周期和共振峰参数的提取等。

3.1　语音信号数字化

语音信号数字化之前，必须先进行防混叠滤波及防工频干扰滤波。其中防混叠滤波指滤除高于 1/2 采样频率的信号成分或噪声，使信号带宽限制在某个范围内；否则，如果采样率不满足采样定理，则会产生频谱混叠，此时信号中的高频成分将产生失真；而工频干扰指 50Hz 的电源干扰。由于防混叠和工频干扰滤波器在一个集成块中，实现起来很简便，在这里不再赘述。

3.1.1　语音信号的采样和量化

语音信号是时间和幅度都连续变化的一维模拟信号，要想在计算机中对它进行处理，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。

在语音信号处理中，需要将信号表示成可以处理的函数的形式。对于模拟信号 $x_a(t)$，它表示函数值随着连续时间变量 t 的变化趋势。如果以一定的时间间隔 T 对这样的连续信号取值，则连续信号 $x_a(t)$即变成离散信号 $x(n)=x_a(nT)$，这个过程称为采样，其中两个取样点之间的间隔 T 称为采样周期，它的倒数 F_s 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，并且可以用采样后的信号重构原始信号。实际的信号常有一些低能量的频谱分量超过采样频率的一半，如浊音的频谱超过 4kHz 的分量比其峰值至少要低 40dB；而对于清音，即使超过 8kHz，频率分量也没有显著下降，因此语音信号所占的频率范围可以达到 10kHz 以上。虽然这样，但对语音清晰度有明显影响部分的最高频率为 5.7kHz 左右。CCITT(国际电报电话咨询委员会)提出的 G.711 标准建议采样频率为 8kHz，但一般情况下这只适合电话语音的情况，因为电话语音的频率为 60～3400Hz。在实际的语音信号处理中，采样频率一般为 8～10kHz。有一些系统为了实现更高质量的语音合成，或者使语音识别系统得到更高的识别率，将可处理的语音信号扩展到 7～9kHz，这时的采样频率一般为15～20kHz。表 3-1 给出了采样率对语音识别系统性能的影响。

表 3-1 不同采样率对误识率降低程度的影响

采 样 率	相对误识率的降低程度	采 样 率	相对误识率的降低程度
8kHz	基线系统	16kHz	+10%
11kHz	+10%	22kHz	+0%

在表 3-1 中,将 8kHz 采样率时的系统作为基线系统,当采样率为 11kHz 时,系统的误识率有 10%的降低;继续升高采样率到 16kHz 时,系统的误识率与 11kHz 相比有 10%的降低;当采样率继续增加时,误识率几乎没有降低。因此在一般的识别系统中,采样率最高选择在 16kHz。

图 3-1 的下半部分为一段模拟信号,其上半部分为对应的离散信号。可以看出,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。量化方式有 3 种:零记忆量化、分组量化和序列量化。零记忆量化是每次量化一个模拟采样值,并对所有采样点都使用相同的量化器特性。分组量化是从可能输出组的离散集合中,选出一组输出值,代表一组输入的模拟采样值。序列量化是在分组或非分组的基础上,用一些邻近采样点的信息对采样序列进行量化。

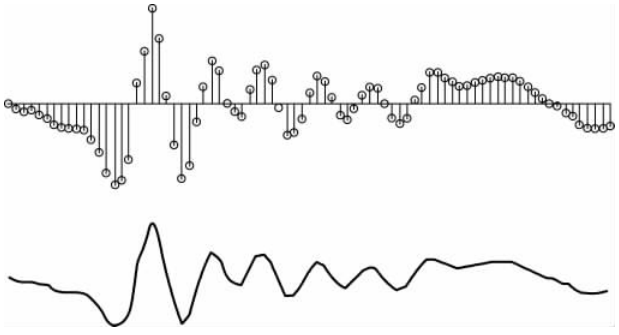

图 3-1 模拟信号和对应的离散信号

零记忆量化是最简单的一种,它的输入-输出特性采用阶梯形函数的形式。图 3-2 给出了两种量化器特性。中点上升量化器的输出没有零电平,在零附近有两个输入区间;正区间产生正输出电平,负区间产生负输出电平。中点水平量化器有零电平输出,它对应于零输入区间。量化范围和电平可以用不同方法选取,但通常都是均匀分布的。

一般量化值都用二进制来表示,如果用 B 个二进制数表示量化值,即量化字长,那么一般将幅度值划分为 2^B 个等分区间。从量化的过程可以看出,信号在经过量化后,一定存在一个量化误差。其定义为

$$e(n) = \hat{x}(n) - x(n) \tag{3-1}$$

其中,$e(n)$为量化误差或噪声;$\hat{x}(n)$为量化后的采样值,即量化器的输出;$x(n)$为未量化的采样值,即量化器的输入。对于上图中的两种量化器,当按 $2x_{\max}=\Delta\times 2^B$ 选定 Δ 和 B 时,量化误差的变化范围为

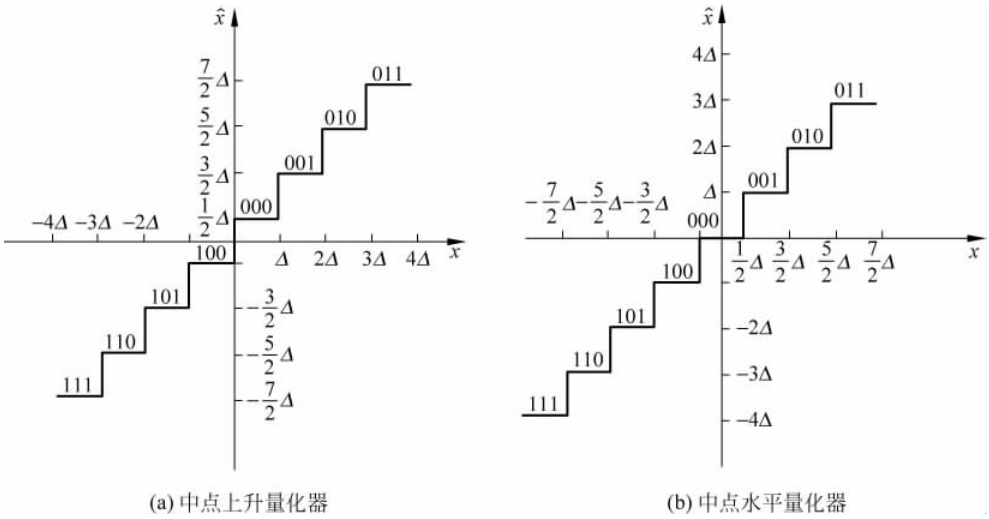

图 3-2　量化器特性

$$-\frac{\Delta}{2} \leqslant e(n) \leqslant \frac{\Delta}{2} \tag{3-2}$$

其中，$x_{\max}$表示信号的峰值，当信号波形的变化足够大或量化间隔 Δ 足够大时，可以证明量化噪声符合具有下列特性的统计模型：①它是一个平稳的白噪声过程；②量化噪声和输入信号相互独立；③量化噪声在量化间隔内均匀分布，即具有等概率密度分布。

若用 σ_x^2 表示输入语音信号序列的方差，σ_e^2 表示噪声序列的方差，则可以证明量化信噪比 SNR(dB)为

$$\mathrm{SNR} = 10\lg\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 6.02B + 4.77 - 20\log\left(\frac{x_{\max}}{\sigma_x}\right) \tag{3-3}$$

假设语音信号的幅度服从拉普拉斯分布，此时信号幅度超过 $4\sigma_x$ 的概率很小，只有0.35%，因而可以取 $x_{\max}=4\sigma_x$。此时式(3-3)变为

$$\mathrm{SNR} = 6.02B - 7.2 \tag{3-4}$$

式(3-4)表明：量化器中每个比特字长对信噪比的贡献大约为 6dB。当量化字长为 7 比特时，信噪比为 35dB。此时量化后的语音质量能满足一般通信系统的要求。然而研究表明，语音波形的动态范围达 55dB，故量化字长应取 10 比特以上。

经过采样和量化过程后，一般还要对语音信号进行一些预加重。由于语音信号的平均功率谱受声门激励和口鼻辐射的影响，高频端大约在 800Hz 以上按着－6dB/倍频程跌落，为此要在预处理中进行预加重。其目的就是提升高频部分，使信号的频谱变得平坦，便于进行频谱分析或声道参数分析。预加重可以在 A/D 变换前，在防混叠滤波之前进行，这样不仅能够进行预加重，而且可以压缩信号的动态范围，有效地提高信噪比。预加重也可以在 A/D 变换之后进行，用具有 6dB/倍频程提升高频特性的预加重数字滤波器实现，预加重滤波器一般是一阶的，形式如下：

$$H(z) = 1 - uz^{-1} \tag{3-5}$$

其中，u 值接近 1，典型的取值为 0.94～0.97。预加重后的信号在分析处理之后，需要进行

去加重处理，即加上－6dB/倍频程下降的频率特性来还原成原来的特性。

一般情况下，如果一个输入信号是若干信号的线性叠加，而其输出是对应的若干输出信号的线性叠加时，则称这样的数字系统为线性系统，否则称其为非线性系统。语音信号处理中常用的非线性系统如表 3-2 所示。

表 3-2 语音信号处理中常用的非线性系统

非线性系统	表 达 式
$(2N+1)$的中值滤波	$y(n)=\text{median}\{x(n-N),\cdots,x(n),\cdots,x(n+N)\}$
全波整流	$y(n)=\lvert x(n)\rvert$
半波整流	$y(n)=\begin{cases}x(n), & x(n)\geqslant 0\\ 0, & x(n)<0\end{cases}$
频率调制	$y(n)=A\cos(\omega_0+\Delta\omega x(n))n$
硬限制器(Hard-Limiter)	$y(n)=\begin{cases}A, & x(n)\geqslant A\\ x(n), & \lvert x(n)\rvert<A\\ -A, & x(n)\leqslant -A\end{cases}$

对于系统的表示，除线性系统和非线性系统外，还可以根据系统参数是否随时间变化分为时不变系统和时变系统。

3.1.2 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，这是由于人在发音时声道一直处于变化状态，因此实际上的语音信号产生系统可以近似看作线性时变系统。为了能用传统的方法对语音信号进行分析，假设语音信号在 10～30ms 短时间内是平稳的。后面的所有分析都是在语音信号短时平稳这个假设条件下进行的。

为了得到短时的语音信号，要对语音信号进行如式(3-6)所示的加窗操作。窗函数平滑地在语音信号上滑动，将语音信号分成帧。分帧可以连续，也可以采用交叠分段的方法，交叠部分称为帧移，一般为窗长的一半。

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。在选择窗函数时，一般有两个问题要考虑。

1. 窗函数形式

窗函数可以选用矩形窗，即

$$w(n)=\begin{cases}1, & 0\leqslant n\leqslant N-1\\ 0, & \text{其他}\end{cases} \tag{3-6}$$

或其他形式的窗函数，如汉明(hamming)窗，即

$$w(n)=\begin{cases}0.54-0.46\cos[2\pi n/(N-1)], & 0\leqslant n\leqslant N-1\\ 0, & \text{其他}\end{cases} \tag{3-7}$$

或汉宁窗，即

$$w(n)=\begin{cases}0.5[1-\cos(2\pi n/(N-1))], & 0\leqslant n\leqslant N-1\\ 0, & \text{其他}\end{cases} \tag{3-8}$$

其中，N 为窗口长度。

这两种窗函数可以统一定义为

$$w(n)=\begin{cases}(1-\alpha)-\alpha\cos[2\pi n/(N-1)], & 0\leqslant n\leqslant N-1\\ 0, & \text{其他}\end{cases} \tag{3-9}$$

其中，汉明窗对应的 $\alpha=0.46$，汉宁窗对应的 $\alpha=0.5$。

虽然这些窗函数的频率响应都具有低通的特性，但不同的窗口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口形状进行比较。

矩形窗在窗内对所有的采样点给以同等的加权，矩形窗函数对应的数字滤波器的单位冲激响应对应的频谱为

$$H(\omega)=\sum_{n=0}^{N-1}\mathrm{e}^{-\mathrm{j}\omega n}=\frac{\sin(\omega N/2)}{\sin(\omega/2)}\mathrm{e}^{-\mathrm{j}\omega(N-1)/2}=A(\omega)\mathrm{e}^{-\mathrm{j}\omega(N-1)/2} \tag{3-10}$$

其中，幅值响应 $A(\omega)$ 是实偶函数，其形状如图 3-3 所示。$A(\omega)$ 穿过横轴的点为 $\omega_k=2\pi k/N$，第一个零值所对应的归一化频率为

$$f_1=\frac{1}{N} \tag{3-11}$$

图 3-3(a)中给出了在 $N=51$ 时的矩形窗及其频率响应的对数幅度。需要注意，f_1 对应于矩形窗的低通滤波器的归一化截止频率。51 点汉明窗的频率响应如图 3-3(b)所示。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，即汉明窗的主瓣带宽大约是同样宽度矩形窗带宽的两倍。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。

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，应用范围也最为广泛。

2. 窗函数长度

不论什么样的窗口，窗的长度对能否反映语音信号的幅度变化起决定性作用。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缓慢地变化，因而也就不能充分地反映波形变化的细节；反之，如果 N 特别小，即等于或小于一个基音周期的量级，则信号的能量将按照信号波形的细微状况而很快地起伏。但如果 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，因此窗口的长度要选择合适。窗的衰减基本上与窗的持续时间无关，因此当改变宽度 N 时，只会使带宽发生变化。

前面的窗口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，基音周期的持续时间会从高音调(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(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，N 选择在 100～200 量级(10～20ms 持续时间)是合适的。

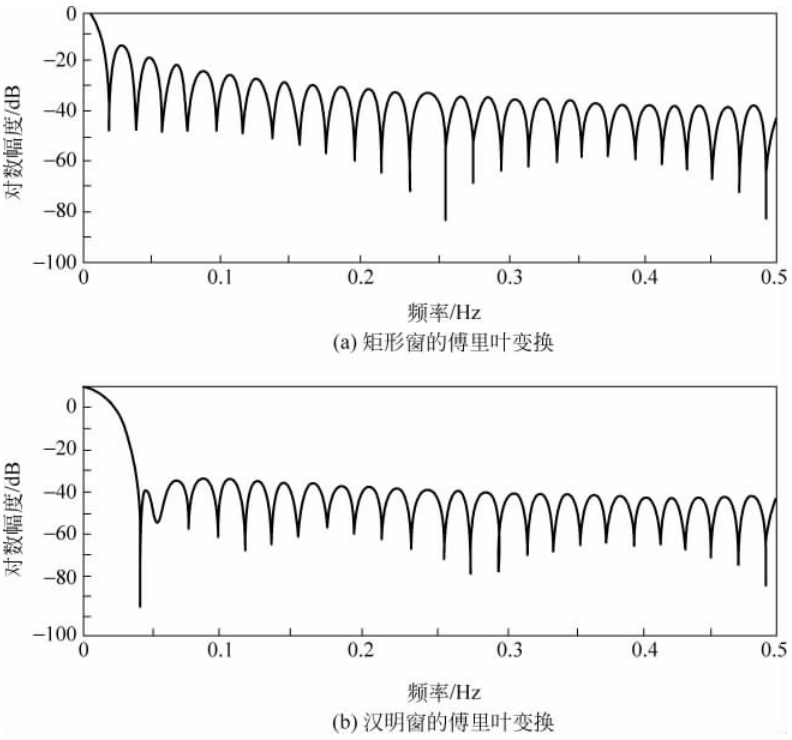

图 3-3 矩形窗和汉明窗的傅里叶变换

3.2 语音信号的时域分析

对信号分析最自然、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。在这一节中主要对这些时域的特征及它们的具体应用加以介绍。

典型的语音信号特性是随着时间变化而变化的。例如，浊音和清音之间激励的改变，会使信号峰值幅度有很大的变化；在浊音范围内基频有相当大的变化。在一个语音信号的波形图中，这些变化十分明显，所以要求能用简单的时域处理技术对这样的信号特征给以有效的描述。

3.2.1 短时能量分析

语音信号的能量随着时间变化比较明显，一般清音部分的能量比浊音的能量小得多。语音信号的短时能量分析给出了反映这些幅度变化的一个合适的描述方法。对于信号$\{x(n)\}$，短时能量的定义如下：

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) = x^2(n) * h(n) \quad (3\text{-}12)$$

其中，$h(n)=w^2(n)$，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，短时能量可以看作语音信号的平方经过一个线性滤波器的输出，该线性滤波器的单位冲激响

应为$h(n)$，如图 3-4 所示。

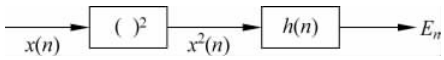

图 3-4　短时能量的方块图表示

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。为了反映窗函数选择对短时能量的影响，假设式(3-12)中的$h(n)$非常长，且为恒定幅度，那么 E_n 随时间的变化将很小，这样的窗就等效为很窄的低通滤波器。很明显，我们要求的是对语音信号进行低通滤波，但还不是很窄的低通滤波，至少短时能量应能反映语音信号的幅度变化。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。但是，太窄的窗将得不到平滑的能量函数。并且窗函数的形状和长短直接影响着短时能量的性质。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，短时能量可表示为

$$E_n = \sum_{m=n}^{n+N-1} x_w^2(m) \tag{3-13}$$

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以区分清音和浊音，因为浊音的能量要比清音的能量大得多；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，以及连字的分界等。在语音识别系统中，短时能量一般也作为特征中的一维参数来表示语音信号的能量大小和超音段信息。

短时能量由于是对信号进行平方运算，因而人为增加了高低信号之间的差距，在一些应用场合不太适用。解决这个问题的简单方法是采用短时平均幅值来表示能量的变化，其公式为

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)|\, w(n-m) = \sum_{m=n}^{n+N-1} |x_w(m)| \tag{3-14}$$

这里用加窗后信号的绝对值之和代替平方和，使运算进一步简化。短时平均幅值的实现如图 3-5 所示。

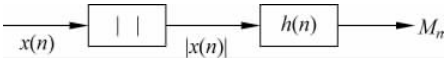

图 3-5　短时平均幅度的方块图

3.2.2　短时平均过零率

短时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，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的频率除以两倍的采样频率，而采样频率是固定的，因此过零率在一定程度上可以反映出频率的信息。语音信号不是简单的正弦序列，所以平均过零率的表示方法就不那么确切。然而短时平均过零率仍然可以在一定程度上反映其频谱性质，可以通过短时平均过零率获得谱特性的一种粗略估计。短时平均过零率的公式为

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]|\, w(n-m)$$

$$= \frac{1}{2}\sum_{m=n}^{n+N-1} \left| \operatorname{sgn}[x_w(m)] - \operatorname{sgn}[x_w(m-1)] \right| \tag{3-15}$$

式中，sgn[·]是符号函数，即

$$\operatorname{sgn}[x(n)] = \begin{cases} 1, & x(n) \geqslant 0 \\ -1, & x(n) < 0 \end{cases} \tag{3-16}$$

图 3-6 给出了短时平均过零率的计算过程。可以看出，首先对语音信号序列 $x(n)$进行成对处理，检查是否有过零现象，若有符号变化，则表示有一次过零现象；然后进行一阶差分计算，取绝对值；最后进行低通滤波。

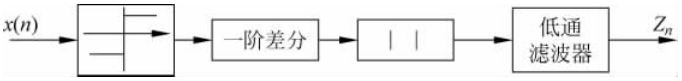

图 3-6 短时平均过零率的计算

短时平均过零率可以用于语音信号分析。在发浊音时，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，在发清音时声带不振动，声道的某部分受到阻塞产生类白噪声的激励，该激励通过声道后能量集中在比浊音时更高的频率范围内。因此，浊音时的能量集中于低频段，而清音的能量集中在高频段。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，一般具有较低的过零率，而在清音段具有较高的过零率，这样可以用短时平均过零率来初步判断清音和浊音。然而这种高低仅是相对而言的，没有精确的数值关系。

另外，可以将短时平均过零率和短时能量结合起来判断语音起止点的位置，即进行端点检测。在背景噪声较小的情况下，短时能量比较准确，但当背景噪声较大时，短时平均过零率可以获得较好的检测效果。因此，一般的识别系统，其前端的端点检测过程都是将这两个参数结合用于检测语音是否真的开始。短时平均过零率的另一个用途是作为语音频域分析的一个中间步骤。方法是不用窗口型的低通滤波器来处理过零，而改用多通道的带通滤波器，这时的输出就是频域的短时平均过零率，如果加上用带通滤波器的短时能量的输出，就可以得到语音信号的频域分析结果。

从上面定义出发计算的短时平均过零率容易受到低频的干扰。解决这个问题的一种方法是对上述定义做一个简单的修改，即设立一个门限 T，将过零率的含义修改为跨过正负门限的次数，如图 3-7 所示。

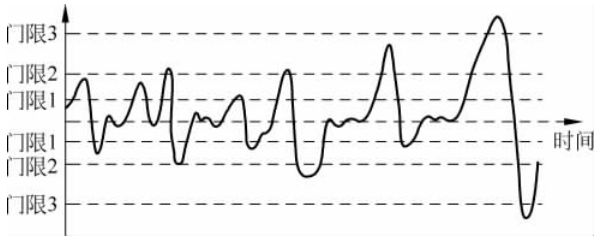

图 3-7 门限短时平均过零率

于是，有

$$Z_n = \frac{1}{2}\sum_{m=-\infty}^{\infty}\{|\ \mathrm{sgn}[x(m)-T]-\mathrm{sgn}[x(m-1)-T]\ | + |\ \mathrm{sgn}[x(m)+T]-\mathrm{sgn}[x(m-1)+T]\ |\}w(n-m) \tag{3-17}$$

这样计算的短时平均过零率就有一定的抗干扰能力。即使存在小的随机噪声，只要它不超过正、负门限所构成的带，就不会产生虚假过零率。在语音识别前端检测时还可以采用多门限过零率，可进一步改善检测效果。

3.2.3　短时自相关函数和短时平均幅度差函数

1. 自相关函数

一般情况下，相关函数用于测定两个信号在时域内的相似程度，可以分为互相关函数和自相关函数。互相关函数主要研究两个信号之间的相关性，如果两个信号完全不同、相互独立，那么互相关函数接近于零；如果两个信号的波形相同，则互相关函数会在超前和滞后处出现峰值，可据此求出两个信号之间的相似程度。自相关函数主要用于研究信号本身的同步性、周期性。这里主要讨论自相关函数的性质及应用。

对于离散的语音数字信号 $x(n)$，它的自相关函数的定义如下：

$$R(k) = \sum_{m=-\infty}^{+\infty} x(m)x(m+k) \tag{3-18}$$

如果信号是随机的或周期的，这时的定义为

$$R(k) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{m=-N}^{N} x(m)x(m+k) \tag{3-19}$$

式(3-18)和式(3-19)表示一个信号和延迟 k 点后的该信号本身的相似程度。在任何一种情况下，信号的自相关函数都是描述信号特性的一种方便的方法。它具有很多性质：

(1) 如果信号 $x(n)$具有周期性，那么它的自相关函数也具有周期性，并且周期与信号 $x(n)$的周期相同；

(2) 自相关函数是一个偶函数，即 $R(k)=R(-k)$；

(3) 当 $k=0$ 时，自相关函数具有最大值，即信号和自己本身的自相关性最大。并且这时的自相关函数值是确定信号的能量或随机信号的平均功率。

从这些性质可以看到，自相关函数相当于一个特殊情况下的能量；而更为重要的是，自相关函数提供了一种获取周期性信号周期的方法。可以看出，在周期信号周期的整数倍上，它的自相关函数可以达到最大值。即可以不用考虑信号的起始时间，而从自相关函数的第一个最大值的位置来估计其周期，这个性质使自相关函数成为估计各种信号周期的一个依据。因此，将自相关函数的定义用到语音信号处理上，以获得其短时自相关函数的表示是十分重要的；这就是下面将介绍的短时自相关函数。

2. 短时自相关函数

短时自相关函数是在前面自相关函数的基础上将信号加窗获得的，即

$$\begin{aligned} R_n(k) &= \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-(m+k)) \\ &= \sum_{m=n}^{n+N-k-1} x_w(m)x_w(m+k) \end{aligned} \tag{3-20}$$

式中，n 表示窗函数是从第 n 点开始加入。通过上述对自相关函数的分析易于证明，$R_n(k)$ 是偶函数，即 $R_n(k)=R_n(-k)$；$R_n(k)$ 在 $k=0$ 时具有最大值，并且 $R_n(0)$ 等于加窗语音信号的能量。

如果定义

$$h_k(n) = w(n)w(n-k) \tag{3-21}$$

那么式(3-20)可以写为

$$R_n(k) = \sum_{m=-\infty}^{+\infty} x(m)x(m-k)h_k(n-m) \tag{3-22}$$

该式表明，序列 $x(n)x(n-k)$ 经过一个冲激响应为 $h_k(n)$ 的滤波器滤波后得到上述的自相关函数，将其用图 3-8 表示如下。

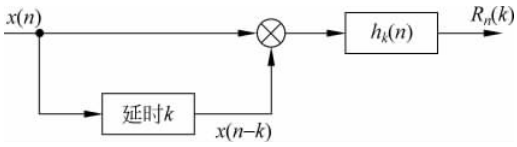

图 3-8　短时自相关函数的计算

如果 $x(n)$ 是一个浊音性的周期信号，那么从自相关函数的性质可知，其短时自相关函数也是呈现出明显的周期性，并且它的周期与原信号本身的周期相同。相反，清音是接近于随机噪声，其短时自相关函数不具有周期性，并随着 k 的增大而迅速减小。因此可以利用这一特点决定一个浊音的基音周期。

图 3-9 给出了三个自相关函数的例子，这是在 $N=401$ 时用 10kHz 采样频率获得的语音计算的自相关函数，并分别计算了滞后为 $0\leqslant k\leqslant 250$ 时的自相关值。前两种情况是对浊音语音段，而第三种情况是对一个清音段。由图 3-9(a)、图 3-9(b)可见，对应于浊音语音的自相关函数，具有一定的周期性。在相隔一定的采样后，自相关函数达到最大值。在图 3-9(c)上自相关函数没有很强的周期峰值，表明在信号中缺乏周期性，这种清音语音的自相关函数有一个类似噪声的波形，有点像语音信号本身。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。在图 3-9(a)中，峰值约出现在 72 的倍数上，由此估计出该浊音的基音周期为 7.2ms 或为 140Hz 左右的基频。在图 3-9(b)中，第一个最大值出现在第 58 个采样的倍数上，它表明平均的基音周期约为 5.8ms。

在语音信号处理中，计算自相关函数所用的窗口长度与计算短时能量时的情况略有不同。这里，N 值至少要大于基音周期的两倍，否则将找不到除 $R(0)$ 外最近的一个最大值点。另一方面，N 值也要尽可能地小，因为语音信号的特性是变化的，N 过大将影响短时性。由于语音信号的最小基频为 80Hz，因而其最大周期为 12.5ms，两倍周期为 25ms，所以 10kHz 采样时窗宽 N 为 250 个采样点。因此，当用自相关函数估算基音周期时，N 不应小于 250。由于基音周期的范围很宽，所以应使窗宽匹配于预期的基音周期。对基音周期较长的信号，使用较窄的窗将得不到预期的基音周期；而对基音周期较短的信号，使用较宽的窗，自相关函数将对许多个基音周期做平均计算，这是不必要的。为此，可采用基于基音周期的自适应窗口长度法，但是这种方法比较复杂。为了解决这个问题，可用“修正的短时自相关函数”来

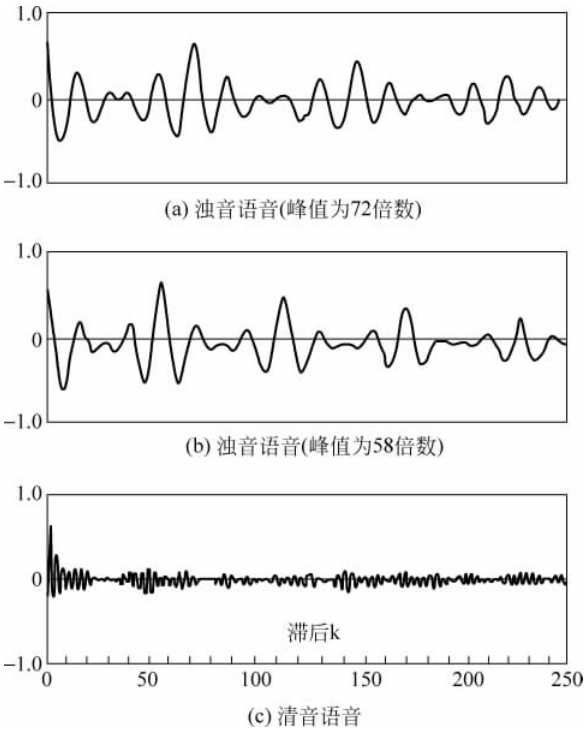

图 3-9　三种自相关函数

代替短时自相关函数。

修正的短时自相关函数定义为

$$\hat{R}_n(k) = \sum_{m=-\infty}^{\infty} x(m)w_1(n-m)x(m+k)w_2(n-m-k) \tag{3-23}$$

或

$$\hat{R}_n(k) = \sum_{m=-\infty}^{\infty} x(n+m)w_1'(m)x(n+m+k)w_2'(m+k) \tag{3-24}$$

与上面公式相比，不同的是两个窗函数用了不同的长度。可以选取 $w_2'(n)$ 使其包括 $w_1'(n)$ 的非零间隔以外的采样，比如在直角窗时，可以使

$$w_1'(m) = \begin{cases} 1, & 0 \leqslant m \leqslant N-1 \\ 0, & \text{其他} \end{cases} \tag{3-25}$$

$$w_2'(m) = \begin{cases} 1, & 0 \leqslant m \leqslant N-1+k \\ 0, & \text{其他} \end{cases} \tag{3-26}$$

因此，修正自相关函数可以写为

$$\hat{R}_n(k) = \sum_{m=0}^{N-1} x(n+m)x(n+m+k) \tag{3-27}$$

式中，k 是最大的延迟点数。

修正短时自相关函数和短时自相关函数计算数据之间的差别如图 3-10 所示。其中图 3-10(a)表示一个语音波形；图 3-10(b)表示由一个矩形窗选取的 N 个采样点；图 3-10(c)

表示 $N+K$ 长度的矩形窗选取的采样点。严格地说，修正自相关函数是两个不同的有限语音段 $x(n+m)w_1'(m)$ 和 $x(n+m)w_2'(m)$ 的互相关函数。因而，$\hat{R}_n(k)$ 具有互相关函数的特性，而不再是一个自相关函数，例如 $\hat{R}_n(k)\neq\hat{R}_n(-k)$。然而 $\hat{R}_n(k)$ 在周期信号的周期倍数上有峰值，所以与 $\hat{R}_n(0)$ 最近的第二个最大值点仍表示基音周期的位置。

(a) 语音波形

(b) 由矩形窗选取的N个采样点

(c) 长度为$(N+k)$的矩形窗选取的采样点

图 3-10　修正短时自相关函数计算中两个不同长度的短时信号说明

3. 短时平均幅度差函数

短时自相关函数是语音信号时域分析的重要参数，但是计算短时自相关函数需要很大的计算量，其原因是乘法运算所需的时间较长。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函数有类似作用的参量，即短时平均幅度差函数。它是基于这样一个想法，对于一个周期为 P 的单纯的周期信号做差分，即

$$d(n)=x(n)-x(n-k) \tag{3-28}$$

则在 $k=0,\pm P,\pm 2P,\cdots$ 时，式(3-28)将为零。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

$$\gamma_n(k)=\sum_{m=n}^{n+N-k-1}|x_w(m+k)-x_w(m)| \tag{3-29}$$

对于周期性的 $x(n)$，$\gamma_n(k)$ 也呈现周期性。与 $R_n(k)$ 相反的是，在周期的各整数倍点上 $\gamma_n(k)$ 具有的是谷值，而不是峰值。因此在浊音语音的基音周期上，$\gamma_n(k)$ 会急速下降，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也可以用于基音周期的检测，而且计算上比短时自相关方法更为简单。

3.2.4　端点检测和语音分割

在许多语音信号处理任务中需要判断一段输入信号中哪些是语音段，哪些是无声段。例如在语音识别中，正确地判定输入语音的起点、终点对于提高识别率往往是非常重要的。在一些语音识别或低速语音编解码器应用中，对于已经判别为语音段的部分，还需要进一步判断清音和浊音。这些问题可以称为有声/无声判决，以及更细致的无声(S)/清音(U)/浊音(V)判决。

能够实现这些判决的依据在于，不同性质语音的各种短时参数具有不同的概率密度函数，以及相邻的若干帧语音应具有一致的语音特性，它们不会在 S、U、V 之间随机地跳来跳去。

在孤立词语音识别系统中，需要正确判断每个输入语音的起点和终点，利用短时平均幅度参数 M 和短时平均过零率 Z 可以做到这一点。首先，根据浊音情况下的短时平均幅度参数的概率密度函数 $P(M|V)$确定一个阈值参数 M_H，M_H 值一般定得较高。当一帧输入信号的短时平均幅度参数超过 M_H 时，可以判定该帧语音信号不是无声，而有相当大的可能是浊音。根据 M_H 可判定输入语音的前后两个点 A_1 和 A_2。在 A_1 和 A_2 之间的部分肯定是语音段，但语音的精确起点、终点还要在 A_1 之前和 A_2 之后仔细查找，如图 3-11 所示。

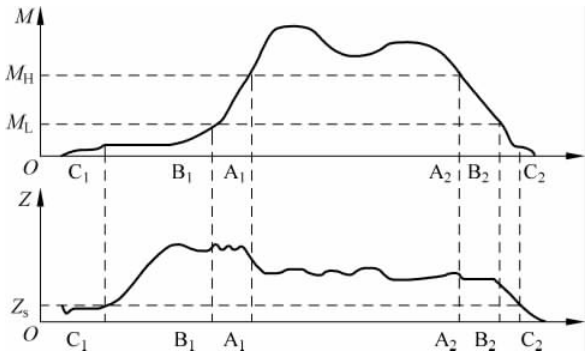

图 3-11　利用短时平均幅度和短时平均过零率判定语音的起点和终点

为此，再设定一个较低的阈值参数 M_L，由 A_1 点向前找，当短时平均幅度由大到小减至 M_L 时，可以确定点 B_1。类似地，可以由 A_2 点向后找，确定 B_2 点。在 B_1 和 B_2 之间仍能肯定是语音段。然后由 B_1 向前和 B_2 向后，利用短时平均过零率进行搜索。根据无声情况下的短时平均过零率，设置一个参数 Z_s，如果由 B_1 向前搜索，短时平均过零率大于 Z_s 的 3 倍，则认为这些信号仍属于语音段，直到短时平均过零率下降到低于 3 倍的 Z_s，这时的点 C_1 就是语音的精确的起点。对于终点做类似的处理，可以确定终点 C_2。采用短时平均过零率的原因在于，点 B_1 以前可能是一段清辅音，它的能量相当弱，依靠能量不可能将它们与无声段分开。而对于清辅音来说，它们的过零率明显高于无声段，因而能用这个参数将二者区分开来。

研究结果表明，利用短时平均过零率来区分无声和清音在有些情况下不是很可靠。由于清音的强度会比无声段高一些，将门限提高一些对于清音的影响不大，但在没有背景噪声的情况下，无声段将不会穿越这一提高的电平，因而可以正确地区分清音和无声段，因此采

用式(3-17)所示的过零率进行判断更加可靠。

除了上述用短时平均幅值和短时平均过零率来进行清浊音判断之外,还可以在求取基音周期时,利用基音周期存在与否来判断是浊音还是清音。

3.3 语音信号的频域分析

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,对语音信号进行频谱分析,是认识语音信号和处理语音信号的重要方法。所采用的分析方法有很多,下面介绍滤波器组分析方法和傅里叶分析方法。

3.3.1 滤波器组方法

利用一组滤波器来分析语音信号的频谱,是最早采用的频谱分析方法之一。这种方法使用简单、实时性好、受外界环境的影响小,所以至今这一方法仍是频谱分析的常用方法。滤波器组法所用的滤波器可以是模拟滤波器,也可以是数字滤波器。滤波器可以用宽带带通滤波器,也可以用窄带带通滤波器。宽带带通滤波器具有平坦特性,用它可以粗略地求取语音的频谱,其频率分辨率降低,相当于短时处理时窗宽窄的那种情况。使用窄带带通滤波器,其频率分辨率提高,相当于短时处理时窗宽较宽的那种情况。图 3-12 为带通滤波器组法频谱分析原理图。

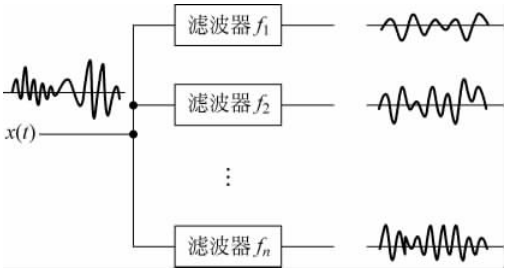

图 3-12 滤波器组法频谱分析原理图

语音信号 $x(t)$输入带通滤波器 $f_1,f_2,\cdots,f_n$,滤波器输出为具有一定频带的中心频率为 $f_1,f_2,\cdots,f_n$ 的信号。图 3-12 中滤波器组的输出为模拟信号,不便于计算机做分析处理。可以将滤波器组的输出经过自适应增量调制器变为二进制脉冲信号,再经过多路开关,变为一串二进制脉冲信号。这种信号可以输入计算机进行各种分析和处理。

3.3.2 傅里叶频谱分析

傅里叶频谱分析是语音信号频域分析中广泛采用的一种方法。它是法国科学家 J. Fourier 在 1807 年为了得到热传导方程的简便解法而提出的。傅里叶变换在电气工程等领域得到了广泛的应用,很多理论研究和应用研究,都把傅里叶变换当作最基本的经典工具来使用。傅里叶频谱分析是分析线性系统和平稳信号稳态特性的强有力的工具,这种以复指数函数为基函数的正交变换,理论上很完善,计算上很方便,概念上易于为人们理解,在语音信号处理上也是一个非常重要的工具。

傅里叶频谱分析的基础是傅里叶变换，用傅里叶变换及其反变换可以求得傅里叶谱、自相关函数、功率谱、倒谱。由于语音信号的特性是随着时间缓慢变化的，由此引出的语音信号短时分析。如同在时域特征分析中用到的一样，这里的傅里叶频谱分析也采用相同的短时分析技术。

信号 $x(n)$ 的短时傅里叶变换定义为

$$X_n(\omega) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)\mathrm{e}^{-\mathrm{j}\omega m} \tag{3-30}$$

式中，$w(n)$ 为窗口函数。

可以从两个角度理解函数 $X_n(\omega)$ 的物理意义：一是当 n 固定时，例如 $n=n_0$，$X_{n_0}(\omega)$ 是将窗函数的起点移至 n_0 处截取信号 $x(n)$，再做傅里叶变换而得到的一个频谱函数。这是直接从频率轴方向来理解的。二是从时间轴方向来理解，当频率固定时，例如 $\omega=\omega_k$，$X_n(\omega_k)$ 可以看作是信号经过一个中心频率为 ω_k 的带通滤波器产生的输出。这是因为窗口函数 $w(n)$ 通常具有低通频率响应，而指数 $\mathrm{e}^{\mathrm{j}n\omega_k}$ 对语音信号 $x(n)$ 有调制的作用，可使频谱产生移位，即将 $x(n)$ 频谱中对应于频率 ω_k 的分量平移到零频。这时的短时傅里叶变换可以理解为如图 3-13 所示的带通滤波器的作用。

图 3-13　从带通滤波器作用理解短时傅里叶变换

在实际计算时，一般用离散傅里叶变换代替连续傅里叶变换，这就需要对信号进行周期性扩展，即把 $x(n)w(n)$ 看成某个周期信号的一个周期，然后对它做离散傅里叶变换，这时得到的是功率谱。值得注意的是，如果窗长为 L，那么 $x(n)w(n)$ 的长度为 L，而 $R_n(k)$ 的长度为 $2L$。如果对 $x(n)w(n)$ 以 L 为周期进行扩展，在自相关域就会出现混叠现象，即这个周期函数的循环相关函数在一个周期中的值就与线性相关 $R_n(k)$ 的值不同，这样得到的功率谱只是真正功率谱的一组欠采样，即 L 个采样值。若想得到功率谱的全部 $2L$ 个值，可以在 $x(n)w(n)$ 之后补充 L 个零，将其扩展成周期为 $2L$ 的信号，并做离散傅里叶变换。这时的循环相关与线性相关是等价的。

图 3-14 给出了几种典型情况下男性元音的短时频谱。可以看出，通过傅里叶变换得到的元音短时频谱中，存在一定数量的峰值。为了说明这个情况，假设 $x_n(m)$ 在窗之外依然保持一种周期性，其周期为 M，对于这样类周期信号的 $x_n(m)$，对应的傅里叶级数的系数为 $X_n(k)$，则其对应的频谱应该是一系列的冲激函数和，即

$$X_n(\omega) = \sum_{k=-\infty}^{\infty} X_n(k)\delta(\omega - 2\pi k/M) \tag{3-31}$$

假设窗函数 $w(m)$ 对应的傅里叶变换表示为

$$W(\omega) = \sum_{m=-\infty}^{\infty} w(m)\mathrm{e}^{-\mathrm{j}\omega m} \tag{3-32}$$

则 $w(n-m)$ 对应的频谱为 $W(\omega)\mathrm{e}^{-\mathrm{j}\omega n}$。因此在时间域信号的乘积 $x(m)w(n-m)$ 在频域上变成卷积关系，即

$$X_n(\omega) = \sum_{k=-\infty}^{\infty} X_n(k)W(\mathrm{e}^{\mathrm{j}(\omega-2\pi k/M)})\mathrm{e}^{\mathrm{j}(\omega-2\pi k/M)n} \tag{3-33}$$

$X_n(\omega)$ 可以看作是幅值由 $X_n(k)$ 控制的若干个窗函数的频谱在每个谐波上平移后的叠

加。这种谐波特性就体现在图 3-14 的窄带峰值上(间隔接近 $2\pi/M$)。

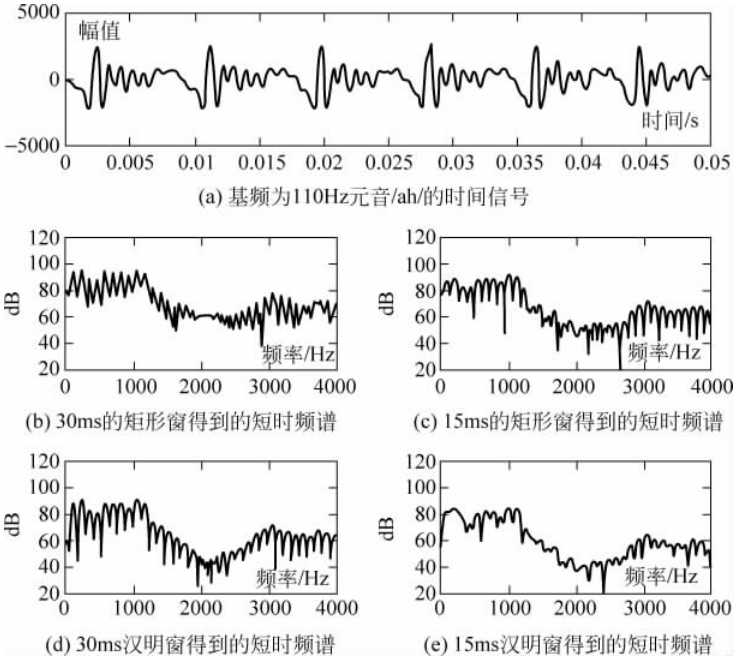

图 3-14 男性元音对应的短时频谱

在窗函数分析中,我们知道对于任一个窗函数都存在旁瓣效应。一般可以对窗函数的频谱近似如下:

$$W(\omega) \approx 0, \quad |\omega - \omega_k| > \lambda \tag{3-34}$$

对于矩形窗函数,窗长为 N,$\lambda = 2\pi/N$。如果 $N \geqslant M$,表明一个窗函数至少包含了一个基音周期,则式(3-34)成立。图 3-14 为基音周期为 $M=71$,采样率为 8kHz 的男声。这里窗长 30ms 对应 $N=240$,窗长 15ms 对应 $N=120$,因此图 3-14(b)和图 3-14(c)均会表现出这种谐波效应,并且窗长越小,对应频谱的主瓣越宽。但对汉明窗,窗长为 N,$\lambda = 4\pi/N$,这就要求一个窗至少包含两个基音周期,即 $N \geqslant 2M$,图 3-14(d)满足这个条件,因此仍然可以看到谐波特性。而对于图 3-14(e),这个条件不再满足,因而谐波特性表现得就不明显。

前面讨论了短时傅里叶变换,从分析中得到语音信号的短时谱 $X_n(\omega)$。下面简要讨论如何由 $X_n(\omega)$来恢复信号 $x(n)$,这就是短时傅里叶反变换。傅里叶变换建立了信号从时域到频域的变换桥梁,而傅里叶反变换则建立了信号从频域到时域的变换桥梁,这两个域之间的变换为一对一映射关系。

我们知道,$X_n(\omega)$可以看作加窗后函数的傅里叶变换,为了实现反变换,将 $X_n(\omega)$进行频率采样,即令 $\omega_k = 2\pi k/L$,则有

$$X_n(\omega_k) = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]e^{-j\omega_k m} \tag{3-35}$$

式中,L 为频率采样点数。

将 $X_n(\omega_k)$在时域 n 上每隔 R 个样本采样,则可令

$$Y_r(\omega_k) = X_{rR}(\omega_k), \quad n = rR, r = 1,2,\cdots \tag{3-36}$$

用这些 $Y_r(\omega_k)$求出其离散傅里叶反变换 $y_r(n)$，即

$$y_r(n) = \frac{1}{L}\sum_{k=0}^{L-1} Y_r(\omega_k)\mathrm{e}^{\mathrm{j}\omega_k n} \tag{3-37}$$

而

$$y(n) = \sum_{r=-\infty}^{+\infty} y_r(n) \tag{3-38}$$

可以证明，$x(n)$和 $y(n)$之间只相差一个比例因子，它们的关系如下：

$$y(n) = x(n)W(0)/R \tag{3-39}$$

即

$$x(n) = \frac{R}{LW(0)}\sum_{r=-\infty}^{+\infty}\sum_{k=0}^{L-1} Y_r(\omega_k)\mathrm{e}^{\mathrm{j}\omega_k n} \tag{3-40}$$

在短时傅里叶变换的基础上，可以得到短时功率谱。短时功率谱实际上是短时傅里叶变换幅度的平方，不难证明，它是信号 $x(n)$的短时自相关函数的傅里叶变换，即

$$P_n(\omega) = |X_n(\omega)|^2 = \sum_{k=-\infty}^{\infty} R_n(k)\mathrm{e}^{\mathrm{j}\omega k} \tag{3-41}$$

式中，$R_n(k)$是前面讨论的自相关函数。

短时功率谱是二维非负的实值函数。用时间作为横坐标，频率作为纵坐标，将短时功率谱的值表示为灰度级所构成的二维图像就是第 2 章中提到的语谱图。下面介绍语谱图中的时间分辨率和频率分辨率。这里分辨率是指对信号所能做出辨别的时域或频域的最小间隔。对时域具有瞬变的信号，希望时域的分辨率要高，即时域的观察间隔尽量短，以保证能观察到该瞬变信号发生的时刻及瞬变的形态。对频域具有两个或多个靠得很近的谱峰信号，希望频域的分辨率要高，即频域的观察间隔尽量短，短到小于两个谱峰的距离，以保证能观察这两个或多个谱峰。

语谱图中的时间分辨率和频率分辨率是由所采用的窗函数来决定的，按照式(3-30)的第一种解释，假定时间固定，对信号乘以窗函数相当于在频域用窗函数的频率响应与信号频谱的卷积。如果窗函数的频率响应 $W(\omega)$的通带宽度为 b，那么语谱图中的频率分辨率的宽度即为 b。即卷积的作用将使任何两个相隔频率小于 b 的谱峰合并为一个单峰。因为对于同一种窗函数而言，其通带宽度与窗长成反比。因此，如果希望频率分辨率高，则窗长应该尽量长一些。

对于时间分辨率，按照式(3-30)的第二种解释，假定频率固定，对信号乘以窗函数的作用，相当于对时间序列 $x(n)\mathrm{e}^{\mathrm{j}n\omega_k}$ 做低通滤波。其输出信号的带宽就是 $w(n)$的带宽 b。根据采样定理，这时只需要以 $2b$ 为采样率就可以充分反映出信号的所有频率成分，可见它所具有的时间分辨率宽度为 $1/(2b)$。因此，如果希望时间分辨率高，则窗长应该尽量取短些。由此可见，时间分辨率和频率分辨率是相互矛盾的，这也是短时傅里叶变换本身固有的缺点。

基于上述分析，在语谱图中分为窄带语谱图和宽带语谱图两种。窄带语谱图用于获得较高的频率分辨率，而宽带语谱图可以获得较高的时间分辨率。

除了前述的短时傅里叶变换频谱和功率谱之外，还有对数功率谱以及倒谱等。其中对

数功率谱就是将功率谱取对数，而倒谱是将功率谱取对数后进行傅里叶反变换，关于倒谱的具体内容在 3.7 节详细介绍。图 3-15 为几种谱之间的关系。

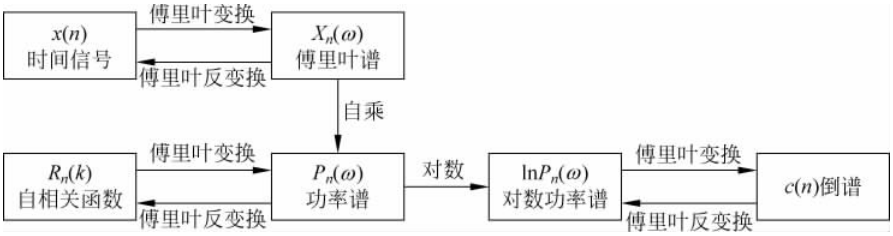

图 3-15　几种基于短时傅里叶变换谱之间的关系

3.4　传统傅里叶变换缺点及时频分析的思想

一般信号都是随着时间的变化而发生变化，要深入理解信号的本质，需要从多个角度研究信号的不同表现方式。时域和频域是观察信号的两种方式，时域分析和频域分析技术也是目前信号处理的主要方法。时域分析方法完全是在时间域中分析信号，时间分辨率理论上可以达到无穷大，但频率分辨率为零，而频域分析方法则相反。一般在频域里分析信号可以得到更多的信息，因此以往人们更重视在频域内对信号加以分析。

自牛顿以来，人们笃信和向往世界的稳定性、规则性、和谐性以及本质上的简单性。傅里叶分析就体现了这种信念。基于傅里叶变换的信号频域表示及其能量的频域分布揭示了信号在频域上的特征。事实上，傅里叶变换是一个强有力的数学工具，它具有重要的物理意义，即信号的傅里叶变换表示信号的频谱。正是傅里叶变换这样重要的物理意义，决定了傅里叶变换在信号分析和信号处理中的独特地位，特别是它可作为平稳信号分析的最重要的工具。然而在实际应用中，所遇到的信号大多数并不是平稳的，至少在观测的全部时间段内它不是平稳的，所以随着应用范围的逐步扩大和理论分析的不断深入，傅里叶变换的局限性就渐渐展示出来。主要表现在如下三个方面。

1. 传统傅里叶变换的时间分辨率为零

传统傅里叶变换的本质在于，它将一个任意的函数表示为一族标准函数的加权和，即正弦函数的加权和。其中的权函数便是原来函数的傅里叶变换。这样就将对原来函数的研究转化为对其权函数，即其傅里叶变换的研究。由于这些正弦函数的频率是固定不变的，并且其波形是无始无终的，因此不难看出，傅里叶分析只适于分析信号组成分量的频率不随时间变化的平稳信号，分析结果也仅能揭示一个信号是由多少个正弦波叠加而成的，以及各正弦波的相对幅度，但不能给出任何有关这些正弦波何时出现与何时消亡的信息。因此，经典的傅里叶分析是一种纯频域分析。理论上频率分辨率可以达到无穷大，但时域内无任何分辨能力，即时域信息完全丧失。傅里叶变换不能反映信号在各个指定时刻的附近所希望的任何频率范围内的频谱信息，这无论在理论上还是在实际中都带来了许多困难和不便。从理论上说，为了用傅里叶变换来研究一个时域信号的频谱特性，就必须获得信号在时域中的全部信息，甚至将来的信息。

2. 传统傅里叶变换基于信号平稳的假设

对于平稳信号，时域分析和频域分析方法都是有效的。传统傅里叶变换的频谱分析是建立在信号平稳假设的基础上。然而，在许多实际应用场合，信号不是平稳的，其统计量是随时间变化的函数。许多天然的和人工的信号，诸如语音、生物医学信号、音乐、雷达和声呐信号、在色散媒质中传播的波、机械振动和动物叫声等都是典型的非平稳信号，其特点是持续时间有限，并且是时变的。对于这种时变信号，必须研究其在时域和频域中的全貌和局部性质，既要能总体上把握信号，又要能深入到信号局部中分析信号的非平稳性，这样才能提取更多的特征信息。这时，只了解信号在时域或频域的全局特性是远远不够的，希望得到的是信号频谱随时间变化的情况。

3. 传统傅里叶变换在全频域范围内分辨率相同

因为一个信号的频率与它的周期成反比，所以在应用中，一个合理的要求是，对于待分析信号的高频信息，其参与分析的信号时间长度应相对较短，以给出精确的高频成分；而对于待分析信号的低频信息，参与分析的信号时间长度应相对较长，以给出一个周期内完整的信息。即要能给出一个对信号进行分析的灵活多变的时间和频率函数，使得由它给出的时域和频域的联合窗口函数宽度具有如下的制约关系：在中心频率高的地方，时间窗自动变窄，而在中心频率低的地方，时间窗应自动变宽。然而，傅里叶变换是一种整体变换，它在整体上将信号分解为不同的频率分量，而对信号的表征要么完全在时域，要么完全在频域。作为频域表示的功率谱，并不能反映出某种频率分量出现在什么时候以及其变化情况。此外，从应用的角度来看，如果一个信号只在某一时刻的一个小的范围内发生变化，那么信号的整个频谱都要受到影响，而频谱的变化从根本上来说又无法标定发生变化的时间位置和发生变化的剧烈程度，即傅里叶变换对信号的局部畸变没有标定和度量的能力。在许多实际的应用中，畸变正是我们所关心的信号在局部范围内的特征，比如对于音乐和语音信号，人们关心的是什么时候演奏什么音符、发出什么音节。

为了分析和处理非平稳信号，人们对傅里叶变换进行了推广，提出并发展了一系列新的信号分析理论。联合时频分析(简称时频分析)就是其中一种重要的方法。它着眼于真实信号组成成分的时变谱特征，将一个一维的时间信号以二维的时间-频率密度函数形式表示出来。时频分析的基本思想是设计时间和频率的联合函数，用该函数同时描述信号在不同时间和频率的能量密度和强度。这种分析方法旨在揭示信号中包含多少频率分量，以及每一分量是怎样随时间变化的。信号的时频表示方法是针对频谱随时间变化的确定性信号和非平稳的随机信号发展起来的。它将一维时域信号 $x(n)$或频域信号 $X(\omega)$映射成为时间频率平面上的二维信号，即使用时间和频率的联合函数来表示信号，这种表示简称为信号的联合时频表示。

3.4.1　信号的时频表示

傅里叶谱和功率谱都是信号变换到频域的一种表示，对于频谱不随时间变化的确定信号及平稳的随机信号，可以用它们进行分析和处理。但当信号的频谱随时间变化时，它不能表示某个时刻信号的频谱分布情况，因此这种分析方法就存在着严重的不足。

针对频谱随时间变化的确定信号和非平稳随机信号，近年来出现了信号的时频域表示方法，如前面 3.3 节中介绍的短时傅里叶变换方法等。其目的是将一维的时间信号 $x(n)$或

频域信号 $X(\omega)$映射成时间-频率平面上的二维信号 $P_x(n,\omega)$。这样,信号的瞬时能量和功率谱可以分别表示为

$$|x(n)|^2 = \int_{-\infty}^{\infty} P_x(n,\omega)\mathrm{d}\omega \tag{3-42}$$

$$|X(\omega)|^2 = \sum_{n=-\infty}^{+\infty} P_x(n,\omega) \tag{3-43}$$

而信号在时频域 $n\in[n_1,n_2]$,$\omega\in[\omega_1,\omega_2]$的能量成分表示为

$$\sum_{n=n_1}^{n_2}\int_{\omega_1}^{\omega_2} P_x(n,\omega)$$

可以根据函数 $P_x(n,\omega)$计算在某一特定时间的频率密度,计算该分布的整体和局部的各阶矩等。然而,在寻求理想的时频表示方法时却遇到了很大的困难。因为理想的 $P_x(n,\omega)$应该表示信号在时间频率点(n,ω)处的能量密度。然而,根据下面即将介绍的不确定性原理,不允许有“某个特定时间和频率处的能量”这一概念,这样理想的 $P_x(n,\omega)$并不存在。因此,只能研究伪能量密度或时频结构,根据不同的要求和不同的性能去逼近理想的时频表示。

人们提出了多种时频表示方法,它们各有优缺点。这些时频表示方法主要有线性时频表示、二次时频表示以及其他形式的时频表示方法。

1. 线性时频表示

这一类时频表示是由傅里叶谱演化而来的,其特点是变换为线性的。由于傅里叶谱具有线性变换的性质,如果信号之间满足线性关系,那么它们的谱函数之间同样满足这样的线性关系,即

$$x(n) = a_1x_1(n) + a_2x_2(n) \tag{3-44}$$

则

$$X(\omega) = a_1X_1(\omega) + a_2X_2(\omega) \tag{3-45}$$

其中,$X(\omega)$,$X_1(\omega)$和 $X_2(\omega)$分别是 $x(n)$,$x_1(n)$和 $x_2(n)$的傅里叶变换;a_1 和 a_2 为常数。因此,由傅里叶谱演化而来的线性时频表示也同样满足这样的线性关系。当 $x_1(n)$和 $x_2(n)$的频谱是随时间变化时,其时频表示 $P_{x_1}(n,\omega)$和 $P_{x_2}(n,\omega)$是线性变换的,则有

$$P_x(n,\omega) = a_1P_{x_1}(n,\omega) + a_2P_{x_2}(n,\omega) \tag{3-46}$$

其中,$P_x(n,\omega)$是 $x(n)$的时频表示。

属于这类的时频表示主要有前面讲述的短时傅里叶变换与 Gabor 变换及小波变换等。其中,短时傅里叶变换和 Gabor 变换是一种加窗的傅里叶变换,使用固定大小的时频网格,时频网格在时频平面上的变化只限于时间平移和频率平移。在短时傅里叶变换和 Gabor 变换这两种时频表示中,窗函数宽度是固定的,其时频分辨率也是固定的,因此只适用于分析具有带宽固定不变的非平稳信号。而实际应用中,常希望在对低频成分分析时,频率的分辨率高一些;对高频成分分析时,时间的分辨率高一些;这就要求窗函数的宽度能随着频率变化而变化。小波变换的时频分析网格的变化除了时间平移外,还有时间和频率轴比例尺度的改变,它使用长宽大小不一的长方形时频分析网格,因而适用于分析具有固定比例带宽的非平稳信号。

2. 二次时频表示

这类时频表示是由能量谱或功率谱演化而来的,其特点是变换为二次的(也称为双线性

的)。能量谱或功率谱具有双线性变换特性,即当信号之间满足式(3-46)的线性关系,则能量谱函数之间为如下的双线性关系:

$$\varepsilon(\omega) = |a_1|^2\varepsilon_1(\omega) + |a_2|^2\varepsilon_2(\omega) + 2\mathrm{Re}[a_1a_2X_1^*(\omega)X_2^*(\omega)] \tag{3-47}$$

其中,$\varepsilon(\omega)$、$\varepsilon_1(\omega)$与 $\varepsilon_2(\omega)$分别为 $x(n)$、$x_1(n)$和 $x_2(n)$的能量谱;$*$ 号表示对信号的频谱取共轭操作。这样,当 $x_1(n)$和 $x_2(n)$的频谱随时间变化时,根据能量谱或功率谱得到的时频表示 $P_{x_1}(n,\omega)$和 $P_{x_2}(n,\omega)$是二次的,则有

$$P_x(n,\omega) = |a_1|^2P_{x_1}(n,\omega) + |a_2|^2P_{x_2}(n,\omega) + 2\mathrm{Re}[a_1a_2P_{x_1x_2}(n,\omega)] \tag{3-48}$$

其中,$P_x(n,\omega)$是 $x(n)$的时频表示;右边最后一项为交叉项或互项;$P_{x_1x_2}(n,\omega)$为 $x_1(n)$和 $x_2(n)$的互时频表示。

维格纳分布是这类时频表示中非常重要的一种。除此之外,还有一些其他二次型能量化的时域表示,可以统一地由 L. Cohen 提出的广义双线性时频表示,即

$$P_x(n,\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{+\infty}\sum_{u=-\infty}^{+\infty}\sum_{\xi=-\infty}^{+\infty}\mathrm{e}^{-\mathrm{j}\xi(n-u)}\varphi(\xi,\tau)x\left(u+\frac{\tau}{2}\right)x^*\left(u-\frac{\tau}{2}\right)\mathrm{e}^{-\mathrm{j}\omega\tau} \tag{3-49}$$

其中,$\varphi(\xi,\tau)$表示核函数,它决定 $P_x(n,\omega)$的特性。

采用不同的核函数,将得到不同的时频分布。对核函数的要求是,希望既能压缩交叉干扰项,又能有好的特性。常用的 Cohen 类广义双线性时频分布有指数分布或称 Choi-Williams 分布、广义指数分布等。

3. 其他时频表示

除了上述线性与二次时频表示外,还有一些其他形式的时频表示,如 Cohen-Posch 类正值分布,L. Stankovic 等人在维格纳分布基础上提出的 L-维格纳分布等。此外,比较重要的还有分数傅里叶变换等。在下面的章节中,将介绍现在应用研究中常见的几种线性时频表示方法:短时傅里叶变换、Gabor 变换、小波变换及它们的联系与区别。

总之,对给定的信号 $x(n)$,人们希望能找到一个二维函数 $P_x(n,\omega)$,它应是人们最关心的两个物理量 n 和 ω 的联合分布函数,可以反映 $x(n)$的能量随时间 n 和频率 ω 变化的形态,同时,又希望 $P_x(n,\omega)$既具有好的时间分辨率,同时又具有好的频率分辨率。但这受到下面将介绍的不确定原理的制约。

3.4.2　不确定原理

在信号分析与信号处理中,信号的“时间中心”及“时间宽度(time-duration)”,以及频率的“频率中心”与“频带宽度(frequency-bandwidth)”是非常重要的概念。它们分别说明信号在时域和频域的中心位置及在两个域内的扩展情况。

如果分别用 $w(n)$和 $W(\omega)$来作为信号的时域和频域表示,则可以用 $\Delta(w)$和 $\Delta(W)$来分别衡量它们的宽度,分别称为有效时域半径和有效频域半径。数值 $2\Delta(w)$和 $2\Delta(W)$称为窗口函数 $w(n)$的有效时宽和有效频宽,而用 $E(w)$和 $E(W)$表示它们的中心。这里中心和半径分别表示为

$$E(w) = \frac{\sum_{n=-\infty}^{+\infty} n\,|w(n)|^2}{\|w\|_2^2} \qquad \Delta(w) = \sqrt{\frac{\sum_{n=-\infty}^{+\infty}(n-E(w))^2\,|w(n)|^2}{\|w\|_2^2}} \tag{3-50}$$

$$E(W)=\frac{\sum_{\omega=-\infty}^{+\infty}\omega\mid W(\mathrm{e}^{\mathrm{j}\omega})\mid^2}{\|W\|_2^2}\quad \Delta(W)=\sqrt{\frac{\sum_{\omega=-\infty}^{\infty}(\omega-E(W))^2\mid W(\mathrm{e}^{\mathrm{j}\omega})\mid^2}{\|W\|_2^2}} \tag{3-51}$$

信号在时间和频率这两个物理量的测量上有一个重要的约束原则,这就是著名的“不确定原理”,或称为“测不准原理”。它的意义是:信号波形在频率轴上的扩张和在时间轴上的扩张不可能同时小于某一界限,即若函数 $w(n)$ 和 $W(\omega)$ 构成一对傅里叶变换,则它们不可能同时都是短宽度的,即

若 $w(n)$ 及其傅里叶变换 $W(\omega)$ 满足窗口函数的条件,则

$$\Delta(w)\Delta(W)\geqslant\frac{1}{2} \tag{3-52}$$

这里等号成立的充分必要条件是 $w(n)$ 为高斯函数,即 $w(n)=A\mathrm{e}^{-\alpha n^2}$。

下面证明这一定理。如果将 $w(n)$ 的导函数的傅里叶变换记为 $W'(\omega)$,那么由傅里叶变换的性质可以得到

$$W'(\omega)=(\mathrm{j}\omega)W(\omega) \tag{3-53}$$

于是,由著名的柯西-施瓦茨(Cauchy-Schwarts)不等式得

$$\begin{aligned}(\Delta(w)\Delta(W))^2&=\frac{1}{\|w\|_2^2}\sum_{n=-\infty}^{+\infty}n^2\mid w(n)\mid^2\cdot\frac{1}{\|W\|_2^2}\sum_{\omega=-\infty}^{+\infty}\omega^2\mid W(\omega)\mid^2\\&=\frac{\sum_{n=-\infty}^{+\infty}n^2\mid w(n)\mid^2\cdot\sum_{\omega=-\infty}^{+\infty}\mid W'(\omega)\mid^2}{\|w\|_2^2\cdot\|W\|_2^2}\\&\geqslant\frac{1}{\|w\|_2^4}\left|\sum_{n=-\infty}^{+\infty}nw(n)w'(n)\right|^2\\&=\frac{1}{\|w\|_2^4}\left(\frac{1}{2}\sum_{n=-\infty}^{+\infty}\mid w(n)\mid^2\right)^2=\frac{1}{4}\end{aligned}$$

所以

$$\Delta(w)\Delta(W)\geqslant\frac{1}{2}$$

在上面推导过程中,等号成立的条件就是 Cauchy-Schwarts 不等式成为等式的条件,最后,通过解微分方程可以得到全部的证明。

不确定原理是信号处理中的一个重要的基本定理,该定理指出,对给定的信号,其时宽与带宽的乘积为一常数。当信号的时宽减小时,其带宽将相应增大,当时宽减到无穷小时,带宽将变成无穷大,例如时域的 δ 函数;反之亦然,例如时域的正弦信号。即信号的时宽与带宽不可能同时趋于无限小,这一基本关系就是前面几节中所讨论过的时间分辨率和频率分辨率的制约关系。在这一基本关系的制约下,人们在竭力探索既能得到好的时间分辨率,又能得到好的频率分辨率的信号分析方法。

3.5 Gabor 变换

传统的傅里叶分析适合于平稳信号处理,它使用的是一种全局的变换。因此,传统的傅里叶分析无法表达信号的时频局域性质。为了分析和处理非平稳信号,人们基于时频分析

思想提出了短时傅里叶变换。3.3 节中从信号处理的角度详细介绍了短时傅里叶变换，本节将从时频分析的角度对短时傅里叶变换进行总结，并将进一步介绍 Gabor 变换。

前面介绍短时傅里叶变换中的“短时”，是直接延续时域分析中对语音的分帧概念而引出的。为了表示信号随时间变化的频谱，采用加窗的技术将信号在时间上分成许多段，然后对每个小段求傅里叶变换，得到对应于不同时刻的信号的频谱，这是短时傅里叶变换的思想。

假定非平稳信号在一个较短的分析窗函数内是平稳（伪平稳）的，移动窗函数，使信号在不同的有限时间宽度内为不同的伪平稳信号，则可以计算出各个不同时刻的功率谱。这些傅里叶变换的集合，就是短时傅里叶变换的结果。显然，这个结果是时间变量和频率变量的二维函数。实际上，在短时傅里叶变换中，对于窗函数有一定的要求。设 $w(n)\in L^2(R)$，即为平方可积空间的函数，而且它的范数不为零，如果 $\sum\limits_{n=-\infty}^{+\infty}|nw(n)|^2<+\infty$，则称 $w(n)$ 是一个窗函数。这时窗函数的中心和半径分别如式(3-50) 和式(3-51) 所示。其中的窗函数有很多种选择，不同的窗函数，对应不同的变换结果。如 3.1.2 节中的矩形窗函数、汉明窗函数以及汉宁窗函数等都是语音信号处理中常用的窗函数。

另外，从时频分析的角度，另一种窗函数——高斯函数也是经常使用的。这时的短时傅里叶变换称为 Gabor 变换。

Gabor 在 1946 年的论文中，为了提取信号的包括时间和频率两方面的局部信息，引入了一个时间局部化的“窗口函数”。所取的窗函数为一个高斯函数，其原因有二：一是高斯函数的傅里叶变换仍为高斯函数，这相当于傅里叶反变换也是用高斯函数加窗的，同时体现了频域的局部化；二是 Gabor 变换作为一般的“窗口函数”具有最佳性，这是在不确定原理明确之后才看出来的，即在时频窗面积最小的意义下，Gabor 变换是最优的窗口傅里叶变换。一般认为只有在 Gabor 变换出现后，才有了真正意义上的时频分析。

对于函数 $x(n)\in L^2(R)$，其 Gabor 变换的定义为

$$G_x(n,\omega)=\sum_{\tau=-\infty}^{+\infty}x(\tau)g_a^*(\tau-n)\mathrm{e}^{-\mathrm{j}\omega\tau} \tag{3-54}$$

式中，$g_a^*(n)=\dfrac{1}{2\sqrt{\pi a}}\exp\left(-\dfrac{n^2}{4a}\right)$ 是高斯函数，a 是大于零的固定常数。

由于 $\sum\limits_{n=-\infty}^{+\infty}g_a(\tau-n)=1$，因此 $\sum\limits_{n=-\infty}^{+\infty}G_x(n,\omega)=X(\omega)$。这表明，信号 $x(n)$ 的 Gabor 变换 $G_x(n,\omega)$ 是对任何 $a>0$ 在时间 $\tau=n$ 附近对 $x(n)$ 傅里叶变换的局部化。对于任意给定 $\omega\in R$，这种局部化完成得很好，达到了对 $X(\omega)$ 的精确分解，从而完整地给出了 $x(n)$ 频谱的局部信息，充分体现了 Gabor 变换在时间域的局部化思想。

对于任意的 $x(n)\in L^2(R)$，它的短时傅里叶变换可写为与 Gabor 变换相似的形式

$$C_x(n,\omega)=\sum_{\tau=-\infty}^{+\infty}x(\tau)w^*(\tau-n)\mathrm{e}^{-\mathrm{j}\omega\tau} \tag{3-55}$$

实际上，如果窗函数 $w(n)$的傅里叶变换也满足窗函数的条件，那么短时傅里叶变换同时也给出了信号 $x(n)$在如下时频窗中的局部信息：

$$[E(w)+n-\Delta(w),E(w)+n+\Delta(w)]\cdot[E(W)+\omega-\Delta(W),E(W)+\omega+\Delta(W)]$$

选定窗口函数 $w(n)$之后,这个时频窗是一条边与坐标轴平行的与(n,ω)无关的矩形,其固定的面积为 $4\Delta(w)\Delta(W)$,该矩形的中心坐标为$(E(w)+n,E(W)+\omega)$。当窗函数的时域中心和频域中心都在原点时,时频窗的中心正好就是参数对(n,ω),这时短时傅里叶变换就真正给出了信号在时间点 n 附近和在频率点 ω 附近,且时频窗为如下形式的时间和频率的局部信息:

$$[n-\Delta(w),n+\Delta(w)]\cdot[\omega-\Delta(W),\omega+\Delta(W)]$$

这也是称它们为时频分析方法的原因所在。

短时傅里叶变换的时频分析能力是用前述时频窗矩形的面积 $4\Delta(w)\Delta(W)$来衡量。在时频窗的形状固定不变时,窗函数面积越小,说明它的时频局部化描述能力越强;窗函数面积越大,说明它的时频局部化描述能力越差。当然,要得到尽量精确的时频局部化描述,自然希望选择使时频窗面积 $4\Delta(w)\Delta(W)$尽量小的窗函数。但是,不确定原理说明这种潜力是有限度的。

对于 Gabor 变换来说,由于高斯函数 $g_a(n)$及其傅里叶变换 $G_a(\omega)$都满足窗函数的要求,可以得到 $g_a(n)$对应的时频窗的面积 $4\Delta(w)\Delta(W)=2$。那么,是否存在比 Gabor 变换所用的高斯函数具有更好的时频局部化描述能力的窗函数呢? 由前面的不确定原理可以知道,当窗函数 $w(n)$及其傅里叶变换都满足窗函数的要求时,$\Delta(w)\Delta(W)\geqslant 1/2$。即 Gabor 变换是具有最小时频窗的短时傅里叶变换,这反映了 Gabor 变换的某种最佳性。当然这里没有考虑到时频窗函数形状的变化与信号时频分析的需要之间的关系。

总之,作为信号分析的工具,短时傅里叶变换和 Gabor 变换发展了傅里叶变换,能够满足信号处理的某些特殊需要。但进一步的研究发现,这两种变换都没有离散的正交基。这决定了它们在进行数值计算时,没有像离散傅里叶变换中 FFT 那样的快速算法,使其应用受到限制;另一方面,当选定窗函数后,对短时傅里叶变换和 Gabor 变换来说,时频窗函数的形状是固定的,它不能随着所分析的信号成分是高频还是低频等信息做相应的变化,而非平稳信号都包含着丰富的频率成分,所以它们对非平稳信号分析能力是有限的。

在对信号做时频分析时,一般对快变的信号,希望它有较高的时间分辨率以观察其快变部分,如尖脉冲等。根据不确定原理,对该信号频域的分辨率必定要下降。由于快变信号对应的是高频信号,对这一类信号采用较高的时间分辨率,就要降低频率分辨率。反之,对慢变信号,由于它对应的是低频信号,所以希望在低频处有较高的频率分辨率,但不可避免地要降低时间分辨率。

下面以矩形窗为例来说明短时傅里叶变换的时频特性。一个宽度为无穷的矩形窗(即直流信号)的傅里叶变换为一 δ 函数,反之亦然。当矩形窗为有限宽时,其傅里叶变换为一函数,即

$$X(\omega)=A\sum_{n=-N}^{+N}\mathrm{e}^{-\mathrm{j}\omega n}=2A\,\frac{\sin\omega(2N+1)/2}{\omega} \tag{3-56}$$

式中,A 是窗函数的高度;N 是其单边宽度。$x(n)$和其频谱 $X(\omega)$如图 3-16(a)和图 3-16(b)所示。

显然,矩形窗的宽度 N 和其频谱主瓣的宽度$\left(-\frac{\pi}{N}\sim\frac{\pi}{N}\right)$成反比。由于矩形窗在信号处理中起到了对信号截短的作用,因此,若信号在时域取得越短,即在时域保持有较高的分辨

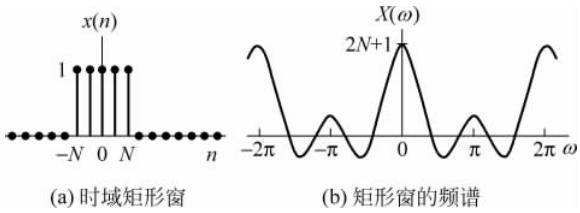

图 3-16　矩形窗及其频谱

率，那么由于 $X(\omega)$的主瓣变宽，因此在频域的分辨率必然会下降。这些体现了短时傅里叶变换中在时域和频域分辨率方面所固有的矛盾。我们希望能用时频分析算法自动适应这一要求。由于短时傅里叶变换窗函数的有效时宽和有效带宽不随(n,ω)的变化而变化，因而它不具备这一自动调节的能力。下面将要讨论的小波变换则具备这一能力。

3.6　小波变换在语音信号分析中的应用

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，它一出现就受到数学界和工程界的极大重视。1984 年法国科学家 J. Molet 在分析地震波的局部特性时，首先使用了小波变换来对信号进行分析，并提出了小波这一术语。所谓小波，就是小的波形，“小”指其具有衰减性，“波”指其波动性，即小波的振幅具有振幅正负相间的振荡形式。小波理论采用多分辨率分析的思想，非均匀地划分时频空间，例如图 3-17 所示的划分方法，它使信号仍能在一组正交基上进行分解，为非平稳信号的分析提供了新途径。

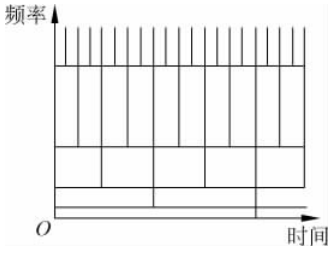

图 3-17　非均匀地划分时间轴和频率轴

3.6.1　小波的数学表示及意义

用数学形式来表述小波，小波就是函数空间 $L^2(R)$中满足下述条件的一个函数或者信号 $\psi(t)$：

$$C_\psi = \int_{R^*} \frac{|\Psi(\omega)|^2}{|\omega|} \mathrm{d}\omega < \infty \tag{3-57}$$

这里，$R^* = R - \{0\}$表示非零实数全体，其中 $\Psi(\omega)$为 $\psi(t)$的频域表示形式。$\psi(t)$称为小波母函数。对于任意的实数对(a,b)，称如下形式的函数为由小波母函数生成的依赖于参数(a,b)的连续小波函数，简称小波。其中参数 a 必须为非零实数。

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3-58}$$

其中，连续性指参数对(a,b) 可以连续取值。若 a,b 不断地变化，可以得到一族函数 $\psi_{a,b}(t)$。对于任意的参数对(a,b)，显然$\int_R \psi_{(a,b)}(t)\mathrm{d}t = 0$。尺度因子 a 的作用是把基本小波 $\psi(t)$ 做伸缩。b 的作用是确定对 $x(t)$ 分析的时间位置，也即时间中心。$\psi_{(a,b)}(t)$ 在 $t = b$ 的附近存在明

显的波动，而且波动的范围大小完全依赖于尺度因子 a 的变化。当 $a=1$ 时，这个范围与原来的小波函数 $\psi(t)$ 的范围是一致的；当 $a>1$ 时，这个范围比原来的小波函数 $\psi(t)$ 的范围大些，小波的波形变得矮宽，而且当 a 变得越来越大时，小波的形状变得越来越宽、越来越矮，整个函数的形状表现出来的变化越来越缓慢；当 $0<a<1$ 时，$\psi_{(a,b)}(t)$ 在 $t=b$ 的附近存在波动的范围比原来的小波母函数 $\psi(t)$ 的波动范围要小，小波的波形变得尖锐而消瘦，当 $a>0$ 且越来越小时，小波的波形渐渐地接近于脉冲函数，整个函数的形状表现出来的变化越来越快。小波函数 $\psi_{(a,b)}(t)$ 随着参数 a 的这种变化规律，决定了小波分析能够对函数和信号进行任意指定点处的任意精细结构的分析，同时，这也决定了小波分析在对非平稳信号进行时频分析时，具有对时频同时局部化的能力。

给定平方可积的信号 $x(t)$，即 $x(t)\in L^2(R)$，则 $x(t)$ 的小波变换定义为

$$W_x(a,b)=\int_R x(t)\psi^*_{(a,b)}(t)\mathrm{d}t=\frac{1}{\sqrt{a}}\int_R x(t)\psi^*\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{3-59}$$

因此，对任意函数 $x(t)$，它的小波变换是一个二元函数，这与傅里叶变换不同。另外，因为小波母函数 $\psi(t)$ 只有在原点附近才会有明显偏离水平轴的波动，在远离原点的地方，函数值将迅速衰减为零，整个波动趋于平静。所以，对于任意的参数对 (a,b)，小波函数 $\psi_{(a,b)}(t)$ 在 $t=b$ 的附近存在明显的波动，远离 $t=b$ 的地方将迅速地衰减到零。因而，从形式上可以看出，小波变换的数值 $W_x(a,b)$ 表明的实质是原来函数 $x(t)$ 在 $t=b$ 附近按照 $\psi_{(a,b)}(t)$ 进行加权平均，体现的是以 $\psi_{(a,b)}(t)$ 为标准快慢的 $x(t)$ 变化情况。这样，参数 b 表示分析的时间中心或时间点，而参数 a 体现的是以 $t=b$ 为中心的附近范围的大小。因此，当 b 固定不变时，小波变换 $W_x(a,b)$ 体现的是原来的函数在 $t=b$ 附近，随着分析和观察的范围逐渐变化时表现出来的变化。

假设小波函数 $\psi(t)$ 及其傅里叶变换 $\Psi(\omega)$ 都满足窗口函数的要求，它们的窗口中心和半径分别记为 $E(\psi)$ 和 $\Delta(\psi)$ 与 $E(\Psi)$ 和 $\Delta(\Psi)$，可以证明对于任意参数对 (a,b)，连续小波 $\psi_{(a,b)}(t)$ 及其傅里叶变换 $\Psi_{(a,b)}(\omega)$ 都满足窗口函数的要求，它们的窗口中心和宽度分别为

$$\begin{cases}E(\psi_{(a,b)})=b+aE(\psi)\\ \Delta(\psi_{(a,b)})=a\Delta(\psi)\end{cases} \tag{3-60}$$

和

$$\begin{cases}E(\Psi_{(a,b)})=E(\Psi)/a\\ \Delta(\Psi_{(a,b)})=a\Delta(\Psi)/a\end{cases} \tag{3-61}$$

因此，对于连续小波 $\psi_{(a,b)}(t)$ 的时间窗为

$$[b+aE(\psi)-a\Delta(\psi),b+aE(\psi)+a\Delta(\psi)]$$

其频率窗为

$$\left[\frac{E(\Psi)}{a}-\frac{\Delta(\Psi)}{a},\quad \frac{E(\Psi)}{a}+\frac{\Delta(\Psi)}{a}\right]$$

因此可以看出，连续小波 $\psi_{(a,b)}(t)$ 的时频窗是时频平面上一个可变的矩形，它的时频窗的面积为

$$2a\Delta(\psi)\times\frac{2\Delta(\Psi)}{a}=4\Delta(\psi)\Delta(\Psi) \tag{3-62}$$

这个面积只与小波的母函数 $\psi(t)$ 有关，而与参数对 (a,b) 毫无关系，但时频窗口的形状随着

参数 a 而发生变化，这是与短时傅里叶变换和 Gabor 变换完全不同的时频分析特性，正是这一点决定了小波变换在信号的时频分析中的特殊作用。

具体地说，对于较小的 $a>0$，这时时间域的窗口宽度 $a\Delta(\psi)$ 随着 a 一起变小，时间窗 $[b-a\Delta(\psi),b+a\Delta(\psi)]$ 变窄（为方便，假定小波的母函数时域中心 $E(\psi)$ 为零），中心频率 $\frac{E(\Psi)}{a}$ 变高，检测到的主要是信号的高频成分。由于高频成分在时间域的特点是变化迅速，因此为了准确检测到在时域中某点处的高频成分，只能利用该点附近很小范围内的观察数据，这必然要求在该点的时间窗比较小，小波变换正好具备了这样的自适应性；反过来，对于较大的 $a>0$，这时时间域的窗口宽度 $a\Delta(\psi)$ 随着 a 一起变大，时间窗 $[b-a\Delta(\psi),b+a\Delta(\psi)]$ 变宽，中心频率 $\frac{E(\Psi)}{a}$ 变低，检测到的主要是信号的低频成分。由于低频成分在时间域的特点是变化缓慢，因此为了完整地检测在时间域中某点的低频成分，必须利用该点附近较大范围内的观测数据，这必然要求在该点的时间窗较大，小波变换恰好具备这种自适应性，这是小波变换作为时频分析方法的独到之处。

3.6.2　小波分析特点

下面从小波变换的恒 Q 性质及时域、频域分辨率，以及与其他变换方法的对比来讨论小波变换的特点，以帮助我们对小波变换有更深入的理解。

若 $\psi(t)$ 的时间中心是 t_0，时宽是 Δ_t，$\Psi(\omega)$ 的频率中心是 ω_0，带宽是 Δ_ω，那么 $\psi\left(\frac{t}{a}\right)$ 的时间中心仍是 t_0，但时宽变成 $a\Delta_t$，$\psi\left(\frac{t}{a}\right)$ 的频谱 $a\Psi(a\omega)$ 的频率中心变为 ω_0/a，带宽变成 Δ_ω/a。这样 $\psi\left(\frac{t}{a}\right)$ 的时宽—带宽积仍是 $\Delta_t\Delta_\omega$，与 a 无关。这一方面说明小波变换的时频关系也受到不确定原理的制约，另一方面，更主要地揭示了小波变换的一个性质，即恒 Q 性质。其中 Q 为母小波 $\psi(t)$ 的品质因数，定义如下：

$$Q=\Delta_\omega/\omega_0=\text{带宽}/\text{中心频率} \tag{3-63}$$

对 $\psi\left(\frac{t}{a}\right)$，其带宽/中心频率为

$$\frac{\Delta_\omega/a}{\omega_0/a}=\Delta_\omega/\omega_0=Q \tag{3-64}$$

因此，不论 a 为何值（$a>0$），$\psi\left(\frac{t}{a}\right)$ 始终保持与 $\psi(t)$ 具有相同的品质因数。恒 Q 性质是小波变换的一个重要性质，也是小波变换区别于其他类型的变换，且被广泛应用的一个重要原因。图 3-18 说明了 $\Psi(\omega)$ 和 $\Psi(a\omega)$ 的带宽及中心频率随 a 变化的情况。

可以看到，正常情况下小波变换如 3-18(a)所示。小波变换在对信号分析时有如下特点：当 a 变大时，对 $x(t)$ 的时域观察范围变宽，频域的观察范围变窄，且分析的中心频率向低频处移动，如图 3-18(b)所示。反之，当 a 变小时，对 $x(t)$ 的时域观察范围变窄，但对 $X(\omega)$ 在频率观察的范围变宽，且观察的中心频率向高频处移动，如图 3-18(c)所示。可以得到在不同尺度下小波变换所分析的时宽、带宽、时间中心和频率中心的关系，如图 3-19 所示。

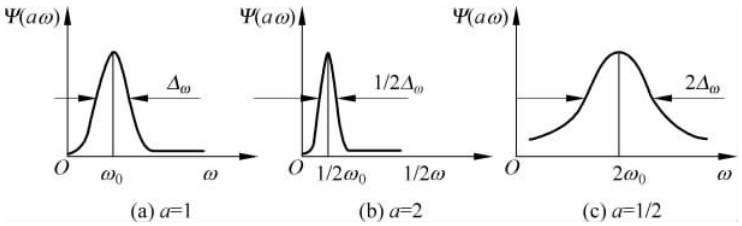

图 3-18 $\Psi(a\omega)$随 a 变化的说明

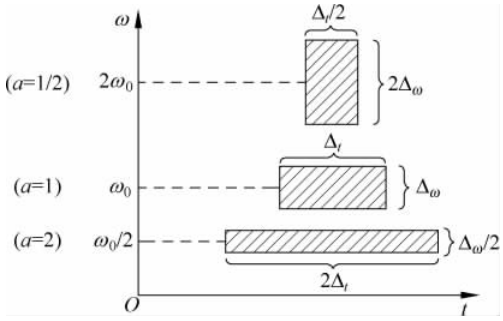

图 3-19 a 取不同值时小波变换对信号分析的时频区间

由于小波变换的恒 Q 性质,因此在不同尺度下,图 3-19 中三个时频分析区间(三个矩形)的面积保持不变。但可以看到,小波变换提供了一个在时频平面上可调的分析窗口。该分析窗口在高频端,如图 3-19 中 $2\omega_0$ 处的频率分辨率不好,矩形窗的频率边变长,但矩形的时间边变短,这表明时域的分辨率增加;反之,在低频端 $\omega_0/2$ 处,频率分辨率变好,而时域分辨率变差。

由小波变换的特点可知,当用较小的 a 对信号做高频分析时,实际上是用高频小波对信号做细致观察;而用较大的 a 对信号做低频分析时,实际上是用低频小波对信号做概貌观察。如上所述,小波变换的这一特点符合对信号做实际分析时的规律。

小波分析是傅里叶分析方法的发展与延拓。它自产生以来,一直与傅里叶分析密切相关。两者相比较主要有以下差别:

(1) 傅里叶变换用到的基本函数只有 $\sin(\omega t)$、$\cos(\omega t)$和 $\exp(j\omega t)$,具有唯一性;小波分析所用到的函数则具有不唯一性,同样一个问题用不同的小波函数进行分析有时结果相差很远。

(2) 在频域中,傅里叶变换具有较好的局部化能力,特别是对于那些频率成分比较简单的确定信号,傅里叶变换可以很容易地把信号表示成各种频率成分叠加和的形式。但在时域中,傅里叶变换没有局部化能力,无法从信号的傅里叶变换中看出原信号在任一时间点附近的形态。

(3) 若用信号通过滤波器来解释,小波变换与短时傅里叶变换的不同之处在于:对短时傅里叶变换来说,带通滤波器的带宽与中心频率无关;相反,小波变换带通滤波器的带宽则正比于中心频率,即小波变换对应的滤波器有一个恒定的相对带宽。

3.6.3　小波变换的多分辨分析

可以用照相机镜头相对被观察景物前后推移的比喻关系来粗略地解释多分辨分析的概念。当尺度 a 较大时，视野宽而分析频率低，可以做概貌的观察；当尺度 a 较小时，视野窄而分析频率高，可以做细节观察，但不同 a 值的品质因数保持不变。这种由粗到细对事物逐级的分析称为多分辨分析，其特性是由信号的自然特征所决定的。一个实际的物理信号不可能在 0～π 的范围内有均匀的频谱。既然信号的能量在不同的频带有不同的分布，在分析时自然需要对它们分别对待。例如，信号在传输过程中需要量化编码，但在有些频段上信号的能量较大，在另一些频段上信号的能量较小。对能量大的频段所对应的信号，应给以较多的比特进行量化编码，而对能量少的频段所对应的信号，可分配较少的比特。这样就可以在保证信号传输质量的前提下，减少所用的比特数。这实际上是对信号进行分层量化。此外，对不同频段所对应的信号还可以采用不同的加权，或者采用不同的去噪处理等。

信号的多分辨率分析，又称信号的多分辨率分解。可以从两个角度引入多分辨分析，即函数空间的划分和理想滤波器组。前者是由 Mallat 首先提出的，数学上比较严谨，结论也比较全面。但是对于具体的信号处理，理想滤波器组则更容易接受，因此我们从理想滤波器组引入多分辨分析的概念。对于函数空间划分方面，只是简要地进行描述。

从理想滤波器组的角度看，多分辨分析实质上是将信号按频带进行分解。信号的分解方法可以是等频带划分，也可以采用一种二进制分解。当信号的采样频率满足采样定理时，归一频带必须限制在 $-\pi\sim+\pi$ 之间。此时可以分别用理想低通滤波器 $H_0(z)$ 和理想高通滤波器 $H_1(z)$ 将其分解成 $0\sim\frac{\pi}{2}$ 的低频部分和 $\frac{\pi}{2}\sim\pi$ 的高频部分，它们分别反映信号的概貌与细节。由于两种滤波器输出的带宽均减半，因此采样频率减半也不至于引起信息的丢失。图 3-20 给出了具体分解的过程。

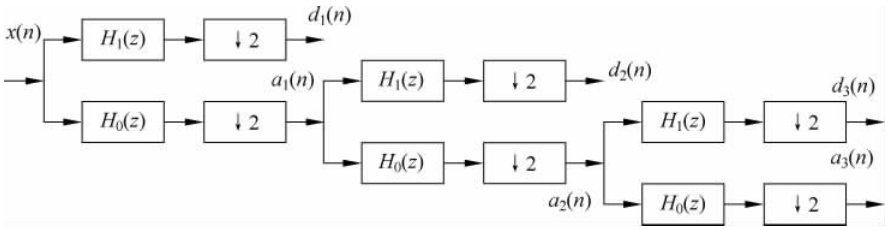

图 3-20　信号二进制分解的实现

如果 $x(n)$ 的带宽在 0～π 之间，采样频率为 f_s，那么经过高通和低通滤波器后，$a_1(n)$ 的带宽在 $0\sim\frac{\pi}{2}$ 之间，$d_1(n)$ 的带宽在 $\frac{\pi}{2}\sim\pi$ 之间，它们均比原信号 $x(n)$ 的带宽(0～π)减小了一半。由此，对 $a_1(n)$ 和 $d_1(n)$ 的采样频率没有必要再用 f_s，仅用 $f_s/2$ 就可以满足采样定理。在上述分解过程中，每一级分解后信号的频带都比前一级减小一半，因此在图 3-20 中每一级都跟随着一个二抽取环节，它表示对每两点数据保存一点，因此采样频率降低了一半。由于 $H_1(z)$ 是高通滤波器，所以其输出 $d_j(n)$ 是每一级的高频信号，称为该级信号的“细节”(detail)，而 $a_j(n)$ 是每一级的低频信号，称为信号的“概貌”或“近似”(approximation)。

从信号的分解过程可以看出，一次次的分解将原信号 $x(n)$ 分成了一个个具有不同频带

的“子带”(subband)信号。若对这些子带信号各自做 DFT,且 DFT 的长度都一样,那么每一个子带信号的频率分辨率是不一样的。对信号 $x(n)$ 的频率分辨率是 f_s/N,对 $a_1(n)$、$d_1(n)$ 的频率分辨率是 $f_s/2N$,提高了一倍,对 $a_2(n)$、$d_2(n)$ 是 $f_s/4N$,对 $a_3(n)$、$d_3(n)$ 的频率分辨率是 $f_s/8N$,这一分析过程是一个由“粗”到“精”的过程。因此,把这一类将原信号按频带分解成一个个子带信号的方法称作“多分辨率分析(或分解)”。

由此可以引出以下概念。

1. 频率空间的划分

如果把原始信号 $x(n)$ 占据的总频带 $0\sim\pi$ 定义为空间 V_0,则经过第一级分解后 V_0 被划分成两个子空间:低频的 $V_1\left(\text{频带 } 0\sim\frac{\pi}{2}\right)$ 和高频的 $W_1\left(\text{频带}\frac{\pi}{2}\sim\pi\right)$。经过第二级分解后 V_1 又被划分成低频 $V_2\left(\text{频带 } 0\sim\frac{\pi}{4}\right)$ 和高频的 $W_2\left(\text{频带}\frac{\pi}{4}\sim\frac{\pi}{2}\right)$,这种子空间分解过程可以记作

$$V_0 = V_1 \oplus W_1, V_1 = V_2 \oplus W_2, \cdots, V_{j-1} = V_j \oplus W_j$$

这些子空间具有逐级包含和逐级替换的特性。

2. 各带通空间具有恒 Q 性

即 W_1 空间的中心频率为 $\frac{3}{4}\pi$,带宽为 $\pi-\frac{\pi}{2}=\frac{\pi}{2}$;$W_2$ 空间的中心频率为 $\frac{3}{8}\pi$,较 W_1 减半,带宽为 $\frac{\pi}{2}-\frac{\pi}{4}=\frac{\pi}{4}$,也较 W_1 减半。可见各 W_j 的品质因数是相同的。

3. 各级滤波器的一致性

各级低通滤波器和高通滤波器是一样的。这是因为前一级输出被二抽取,而滤波器设计是根据归一频率进行的,所谓归一频率是指真实频率与采样间隔的乘积。例如第一级低通滤波器的真实频带是 $0\sim\frac{\pi}{2T_s}$(T_s 是输入的采样间隔),其归一频率则是 $0\sim\frac{\pi}{2}$。第二级低通滤波器的真实频带虽然是 $0\sim\frac{\pi}{4T_s}$,但归一频率仍是 $0\sim\frac{\pi}{2}$,因为第二级输入的采样间隔是 $2T_s$。

从函数空间划分的角度看,在二分的情况下 Mallat 从函数的多分辨率空间分解概念出发,在小波变换与多分辨分析之间建立起联系。如果把平方可积的函数 $x(t)\in L^2(R)$ 看成是某一逐级逼近的极限情况,则每级逼近都是用某一平滑函数对 $x(t)$ 做平滑的结果,只是逐级逼近时平滑函数也做逐级伸缩,即用不同的分辨率来逐级逼近待分析的函数 $x(t)$。对于 V_j 与 W_j 空间,可以找到相应空间的标准正交基,并可以由此构造尺度函数 $\phi(t)$ 与小波函数 $\psi(t)$。其中尺度函数和低通滤波器相对应,而小波函数和高通滤波器相对应。

3.6.4 小波变换在语音处理中的应用

如前所述,小波变换具有很多傅里叶变换无法比拟的性质,使得小波变换在非平稳信号的分析和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,并开展了相关的研究工作,主要包括:利用小波变换对听觉感知系统进行模拟,对语音信号去噪,进行清、浊音判断。

1. 利用小波变换对听觉系统的模拟

听觉系统对声音信号的感知是一系列复杂的转换过程，这些转换大致分为三个阶段：耳蜗滤波器，也就是基底膜完成对信号的分析；毛细胞完成机械振动到点激励的转换；侧抑制网络完成声学谱特征的缩减。对声音信号的分析主要是在基底膜上完成的。基底膜上的振动是以行波方式传递的。频率不同，行波传播的距离也不同，从而不同频率行波的极大值出现在基底膜的不同位置上。频率高的极大值在基底膜的前端，频率低的极大值在其末端，这使得基底膜具有频率分解的能力。此外，对相同的频差，振动频率低时其极大值相距较远，而振动频率高时其极大值相距较近。因此，基底膜对低频的分辨力要高于高频的分辨力。

由于人耳的频率分辨率是非线性的，用传统的线性信号处理方法，如傅里叶变换来模拟人耳基底膜的频率分析特性是比较困难的。可以利用小波变换对频带进行划分，使得其接近于临界频带。使用单纯的小波变换对信号进行处理时，是将整个频带二分，然后保留高频部分，对低频部分继续二分，如此重复下去。这样当频带为 4kHz 时，得到各个子带带宽依次为 2kHz、1kHz、500Hz 和 125Hz，如图 3-21 所示，这与临界频带的划分相去甚远。

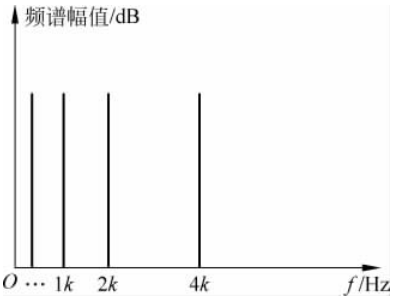

图 3-21　小波变换对频带的划分

为此可以采用广义的小波变换，即把小波变换与小波包变换结合使用，以不完全的小波包变换来对输入信号进行处理。小波包算法有灵活的时频分析能力，可以更好地符合人耳基底膜的频率分析特性。这时对频带的划分如图 3-22 所示。进行小波包变换时阶数最大为 5，当频带宽为 4kHz 时，子带最小宽度为 125Hz，接近最小的临界频带带宽。

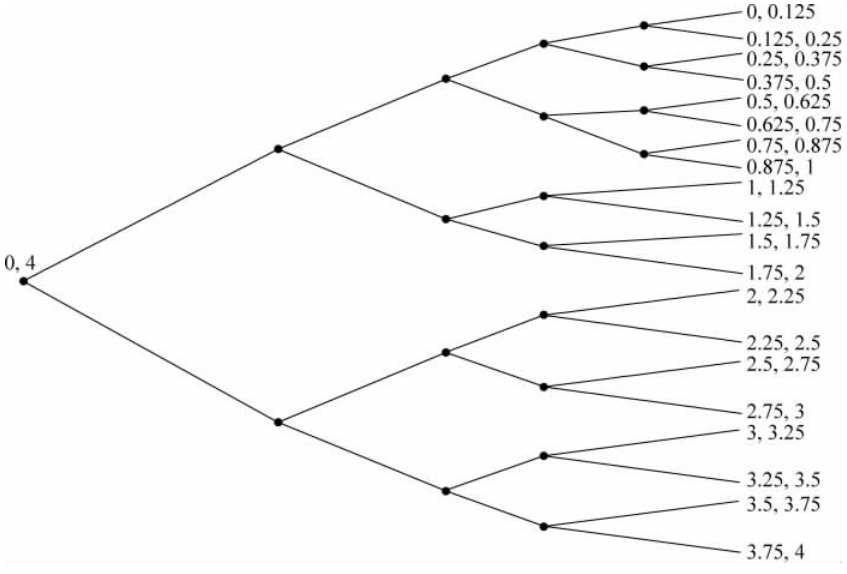

图 3-22　不完全小波包变换对频带的划分

2. 用于随机噪声的去除

传统的基于滤波的噪声去除方法是将被噪声污染的信号通过一个滤波器，滤掉噪声频率成分。但是对于短时瞬态信号、非平稳过程信号、含宽带噪声信号，采用传统方法进行处理有着明显的局限性。小波变换具有时频局部分析的特点，具有传统方法不可比拟的、非常灵活的对奇异特征提取的功能，可在低信噪比的情况下有效地去噪，并检测信号的波形特征。

利用小波变换去噪的基本思想是：根据噪声与信号在各尺度(即各频带)上的小波谱具有不同表现的特点，将噪声小波谱占主导地位的那些尺度上的噪声小波谱分量去掉，这样保留下来的小波谱基本上就是原信号的小波谱，然后再利用小波变换重构算法，重构出原信号。小波变换去噪的关键是如何滤除由噪声产生的小波谱分量。

白噪声信号在小波变换下具有与语音信号不同的特点，这可以由以下的两个定理来体现。

定理 3.1 假设一个信号 $n(t)$是一个方差为 σ^2 的宽平稳白噪声，$\psi(t)$是一个小波函数，则白噪声 $n(t)$的小波变换的期望值为

$$E\{|W_s n(t)|^2\} = \frac{\|\psi\|^2}{s}\sigma^2 \tag{3-65}$$

即 $E\{|W_s n(t)|^2\}$的衰减正比于 $1/s$，随着小波变换尺度的增加，白噪声的小波变换幅值平均减少；即噪声的能量随尺度的增大而迅速减少。

定理 3.2 若白噪声 $n(t)$是高斯白噪声，在尺度 s，其小波变换模的平均密度为

$$d_s = \frac{1}{s\pi}\left(\frac{\|\psi''\|}{\|\psi'\|} + \frac{\|\psi'\|}{\|\psi\|}\right) \tag{3-66}$$

该定理说明白噪声的小波变换模值的平均密度正比于 $1/s$，随着尺度 s 增大，其密度减小。另外，还可以证明高斯白噪声几乎处处奇异。

由上述两个定理可知，随着尺度的增加，白噪声的小波谱将逐渐消失，而有效信号的小波变换在大尺度上仍有清楚的表现。因此，通过观察信号与噪声小波谱模值随尺度增加或减少的演变情况，可以区分白噪声及信号各自产生的变换模值。如果 s 减少，小波变换模幅值急剧增加，则说明这些模值主要由白噪声产生，应该去掉。另外，噪声在不同尺度下的小波变换是高度不相关的；信号的小波变换一般具有很强的相关性，相邻尺度上的局部模极大值几乎出现在相同的位置上，并且有相同的符号。可以利用这点判断小尺度上哪些成分属于有用信号，应予以保留；哪些成分属于噪声，应予以滤除。由于小波基函数的局部支撑特性，能够改变信号在某些点或某些段的值，而不影响到其他部分。这是小波消除噪声比傅里叶变换去除噪声更灵活有效的原因之一。

在去噪时通常采用二进小波，通过分析小波变换的模极大值进行去噪，具体步骤如下：

(1) 带噪信号进行小波变换，提取所有模的极大值，一般最大尺度 J 会小于 4；

(2) 求取阈值 $T_0 = C\dfrac{M}{J}$，其中 M 为最大尺度 $s=2^J$ 上的最大幅值，C 为一个常数；

(3) 在最后一个尺度 J 上，将小波变换后幅值小于阈值 T_0 处的点全部去掉，因为在这些点上噪声的小波变换分量仍有影响；

(4) 将小波变换后的大于阈值的部分求出相应的 α，其中 $\alpha = \log_2\left|\dfrac{W_{2^{j+1}}f(x)}{W_{2^j}f(x)}\right|$，一般取

$j=3$ 或 4，若某点 t 处的 α 小于 0，则令 α 为 0；

(5) 将 1，…，$J-1$ 尺度的小波变换全部去掉，由最后一个尺度的小波变换，按照 $W_{2^j}f(x)=W_{2^{j+1}}f(x)\times 2^{-\alpha}$ 重新构造出 $j=J-1$，…，1 尺度上的小波变换；

(6) 由重建的小波变换经小波反变换恢复去噪后的信号。

3. 用于清音和浊音判断

语音信号小波系数的低频部分描述了信号的轮廓，相当于信号经过低通滤波器的结果；高频部分描述了信号的细节，相当于信号经过高通滤波器的结果。根据语音信号短时平稳的特点，首先对语音信号分帧进行小波变换，将小波域的系数平均分为 4 个频带，计算每个频带的平均能量。如果满足以下条件：①在小波域中的最高频带的能量比其他频带的能量大；②最低频带的能量和最高频带的能量比小于 0.9，则认为这段语音信号为清音。

另外，小波变换还可以用于动态频谱分析。例如，将其用于语音信号分析，看它是否能比传统的语谱图揭示出更多的信息，特别是关于快变语音段的特征；或利用小波变换作为携带信号信息的载体，在语音识别中用作特征提取的手段，而不关心它是否能表示功率谱密度。

3.7　语音信号的同态解卷积

按照语音信号产生的线性模型理论，语音信号是由激励信号与声道响应卷积产生的。在语音信号处理所涉及的各个领域中，根据语音信号求得声门激励信号和声道冲激响应有着非常重要的意义。例如，为了求得语音信号的共振峰，必须知道声道的传递函数。又如，为了判断语音信号是清音还是浊音，以及求得浊音情况下的基音频率，必须知道声门激励序列。要想提取反映声道特性的谱包络，就必须通过解卷积去掉激励信息。

“解卷”，就是将各卷积分量分开。解卷算法可以分成两大类。一类算法称为“参数解卷”，即线性预测分析；另一类算法称为“非参数解卷”，即同态解卷积，对语音信号进行同态分析后，将得到语音信号的倒谱参数，因此同态分析也称为倒谱分析或同态处理。同态处理是一种较好的解卷积的方法，它可以较好地将语音信号中的激励信号和声道响应分离，并且只需要用十几个倒谱系数就能相当好地描述语音信号的声道响应，因而在语音信号处理中占有很重要的位置。本节主要介绍同态处理的基本原理，以及声道响应和激励源的倒谱特性和一些常用的语音特征表示等。

3.7.1　同态信号处理的基本原理

通常的加性信号可以用线性系统来处理，这种系统是满足线性叠加原理的。然而许多客观物理现象中的信号，其中各组成分量的组合，并不是按加法组合原则组合起来的。例如，图像信号、地震信号、通信中的衰落信号、调制信号以及我们所研究的语音信号等，都不是加性信号；而是乘积性组合信号或卷积性组合信号。显然，这样的信号不能用线性系统来处理，而必须用满足该组合规则的非线性系统来处理才行。但是非线性系统分析起来非常困难。同态信号处理法就是设法将非线性问题转化为线性问题来处理的一种方法。按被处理信号来分类，大体上可以分为：乘积同态信号处理和卷积同态信号处理两种。由于语音信号可以看作是声门激励信号与声道响应的卷积结果，所以下面仅讨论卷积同态信号处

理问题。

同态信号处理的一个通用系统构成如图 3-23 所示。其中,符号 * 表示由卷积组合规则组合起来的空间,即该系统的输入和输出信号都是卷积性信号。同态系统的一个最主要理论结果是同态系统分解,分解的目的是用两个特征系统和一个线性系统来代替非线性的同态系统。分解的情形如图 3-23(b)所示。针对语音信号的具体情况,其特征系统和逆特征系统及其运算情况如图 3-23(c)、(d)所示。

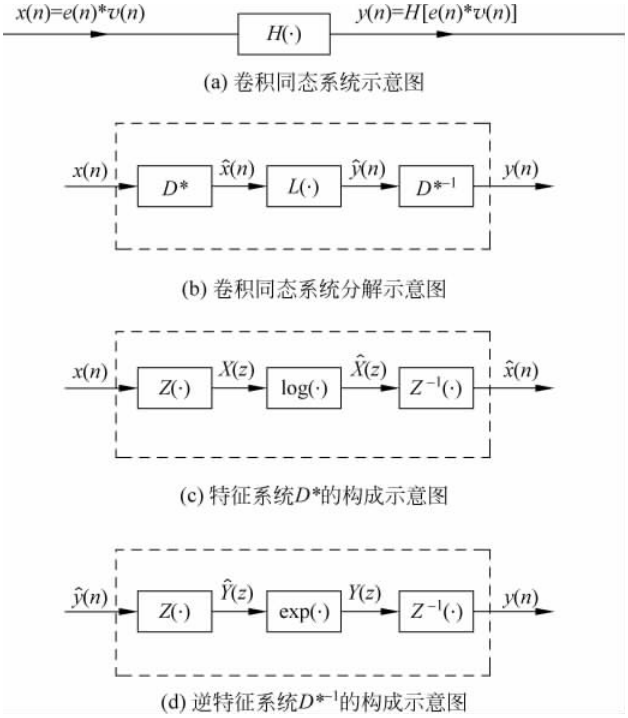

图 3-23 卷积同态系统及其分解和特征系统的构成

假设输入信号是两个信号的卷积,这两个信号 $e(n)$和 $v(n)$分别对应声门激励信号和声道响应序列。特征系统 D^* 的运算是将卷积信号转化为加性信号。它包括三步。第一步是对信号进行 Z 变换,将卷积信号转变为乘积信号,这时得到的就是输入信号的频谱:

$$Z[x(n)] = X(z) = E(z) \times V(z) \tag{3-67}$$

第二步是进行对数运算,将乘积信号变为加性信号:

$$\log X(z) = \log E(z) + \log V(z) = \hat{E}(z) + \hat{V}(z) = \hat{X}(z) \tag{3-68}$$

由于这个信号是加性的对数频谱,使用起来有些不方便,因此常常将它再变回时域信号。所以第三步进行 Z 反变换运算,得到的就是输入语音信号的倒谱(cepstrum):

$$Z^{-1}[\hat{X}(z)] = Z^{-1}[\hat{E}(z) + \hat{V}(z)] = \hat{e}(n) + \hat{v}(n) = \hat{x}(n) \tag{3-69}$$

由于加性信号的 Z 变换或 Z 反变换的结果仍然是加性信号,所以倒谱这种时域信号是可以用线性系统加以处理的。

$L(\cdot)$是在倒谱域对信号进行处理,常见的处理方式是将语音声源信号和声道信号分离。

由于在倒谱域，总可以找到一个 N，当 $n \geqslant N$ 时，声道滤波器的倒谱为零。而在 $n < N$ 时，激励的倒谱接近于零。这样在图 3-24 中，可以通过 $l(n)$形式分别把激励和声道的倒谱信息进行分离。

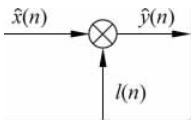

图 3-24　解卷积的倒谱域常见的处理方式

对于图 3-24，为得到声道的倒谱信息，其对应的 $l(n)$如下：

$$l(n) = \begin{cases} 1, & |n| < N \\ 0, & |n| \geqslant N \end{cases} \tag{3-70}$$

同理，为得到激励的倒谱信息，其对应的 $l(n)$的表示如下：

$$l(n) = \begin{cases} 0, & |n| < N \\ 1, & |n| \geqslant N \end{cases} \tag{3-71}$$

经过 $L(\cdot)$处理之后，如果想再恢复为语音信号 $y(n)$，可以用图 3-23(d)所示的逆特征系统运算。显然，它是特征系统的反运算，即将线性系统输出的加性倒谱信号进行 Z 变换，得到线性的对数频谱，然后再进行指数运算转换为输出频谱，这种频谱是一种乘积性信号。最后通过 Z 反变换，就得到卷积性的语音恢复信号。

3.7.2　语音信号的复倒谱

在倒谱域上，可以将信号分为实倒谱信号和复倒谱信号。对于输入信号 $x(n)$，如果其对应的倒谱信号求解如式(3-72)表示，则其对应定义为实倒谱信号。

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |X(\omega)| e^{j\omega n} d\omega \tag{3-72}$$

如果在其倒谱域的求解过程中，不仅考虑信号对应的频谱的模，也考虑其相位，则称其为复倒谱域。这时对应的公式可以表示如下：

$$\hat{x}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log X(\omega) e^{j\omega n} d\omega \tag{3-73}$$

如果采用复倒谱的表示，则需要对复数的频谱信号取对数，这时的对数表示为

$$\hat{X}(\omega) = \log X(e^{j\omega}) = \log |X(\omega)| + j\theta(\omega) \tag{3-74}$$

其中相位为

$$\theta(\omega) = \arg[X(\omega)] \tag{3-75}$$

1. 声门激励信号

除了人们发清音时，声门激励是能量较小、频谱均匀分布的白噪声外；在发浊音时，声门激励是以基音周期为周期的冲激序列：

$$e(n) = \sum_{r=0}^{M} \alpha_r \delta(n - rN_p) \tag{3-76}$$

式中，M 是正整数；α_r 是振幅因子；N_p 为基音周期。这样的冲激序列的 Z 变换为

$$E(z) = \sum_{n=-\infty}^{+\infty}\left[\sum_{r=0}^{M} \alpha_r \delta(n - rN_p)\right] z^{-n} = \sum_{r=0}^{M} \alpha_r z^{-rN_p} \tag{3-77}$$

由式(3-77)可见，$E(z)$是变量 z^{-N_p} 的多项式，而不是 z^{-1} 的多项式。于是，$E(z)$可以表示成形式为$(1-az^{-N_p})$因式的乘积，即

$$E(z) = \alpha_0 \prod_{r=1}^{M} [1 - a_r (z^{N_p})^{-1}] \tag{3-78}$$

通常由于 $a_r = \alpha_r / \alpha_0$ 小于 1，所以将上述公式取对数，并用泰勒公式展开为

$$\widehat{E}(z) = \log E(z) = \log\alpha_0 - \sum_{r=1}^{M} \sum_{k=1}^{+\infty} \frac{a_r^k}{k} (z^{N_p})^{-k}, \quad |z^{N_p}| > |a_r| \tag{3-79}$$

因此，对上式求 Z 的反变换，就可以得到倒谱：

$$\hat{e}(n) = \log\alpha_0 \delta(n) + \sum_{k=1}^{+\infty} \beta_k \delta(n - kN_p) \tag{3-80}$$

式中：

$$\beta_k = -\frac{1}{k} \sum_{r=1}^{M} a_r^k = -\frac{1}{k} \sum_{r=1}^{M} \left(\frac{\alpha_r}{\alpha_0}\right)^k, \quad 1 \leqslant k \leqslant +\infty \tag{3-81}$$

由声门激励的倒谱可以得到如下结论：①一个周期冲激的有限长度序列，其倒谱也是一个周期冲激序列，而且周期长度 N_p 不变，只是长度变成无限长度；②周期冲激序列倒谱的振幅随着 r 值的增大而衰减，并且衰减的速度比原序列要快。

这些特点对语音信号的分析很有用。这意味着除了原点外，可以采用“高时窗”来从语音信号的倒谱中提取浊音信号的倒谱，从而使得用倒谱法提取基音周期成为现实。

声门激励源在浊音时，其倒谱只在 $n = kN_p$ 诸点上不等于零，在其他点上均为零。即声门激励在浊音时，倒谱序列第一个非零点与原点的距离正好为基音周期 N_p。在清音的情况下，声门激励源具有噪声特性，因而这时的倒谱没有明显的峰点，分布范围很宽，从低时域延伸到高时域。利用这个特点可以进行清音和浊音的判断。

2. 声道冲激响应的倒谱

如果用最严格的极零模型来描述声道响应，则该响应序列 $v(n)$ 的 Z 变换有如下的形式：

$$V(z) = |A| \frac{\prod_{k=1}^{m_1} (1 - a_k z^{-1}) \prod_{k=1}^{m_0} (1 - b_k z)}{\prod_{k=1}^{p_1} (1 - c_k z^{-1}) \prod_{k=1}^{p_0} (1 - d_k z)} \tag{3-82}$$

式中，A 是一实数，它是归一化 $V(z)$ 后得到的一个系数。而 $|a_k|$、$|b_k|$、$|c_k|$、$|d_k|$ 的值都小于 1。上式表明，$V(z)$ 具有 m_1 个零点在 Z 平面单位圆内，有 m_0 个零点在 Z 平面单位圆外；有 p_1 个极点在 Z 平面单位圆内，有 p_0 个极点在 Z 平面单位圆外。

将式(3-82)求对数即可得到

$$\hat{V}(z) = \log V(z) = \log|A| + \sum_{k=1}^{m_1} \log(1 - a_k z^{-1}) + \sum_{k=1}^{m_0} \log(1 - b_k z)$$
$$- \sum_{k=1}^{p_1} \log(1 - c_k z^{-1}) - \sum_{k=1}^{p_0} \log(1 - d_k z) \tag{3-83}$$

除了 $\log|A|$ 外，上式所有项都包含 $\log(1 - \alpha z^{-1})$ 和 $\log(1 - \beta z)$ 的形式，这些因式所表示的 Z 变换的收敛域都包括单位圆。由于 $|a_k|$、$|b_k|$、$|c_k|$、$|d_k|$ 都小于 1，所以可以用泰勒展开将上式的后 4 项按下面模式展开：

$$\log(1 - \alpha z^{-1}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n}, \quad |z| > |\alpha| \tag{3-84}$$

$$\log(1-\beta z)=-\sum_{n=1}^{\infty}\frac{\beta^n}{n}z^n,\quad |z|<|\beta^{-1}| \tag{3-85}$$

将上述类型的展开式代入式(3-83),有

$$\hat{V}(z)=\log|A|-\sum_{k=1}^{m_1}\sum_{n=1}^{+\infty}\frac{a_k^n}{n}z^{-n}-\sum_{k=1}^{m_0}\sum_{n=1}^{+\infty}\frac{b_k^n}{n}z^n$$
$$+\sum_{k=1}^{p_1}\sum_{n=1}^{+\infty}\frac{c_k^n}{n}z^{-n}+\sum_{k=1}^{p_0}\sum_{n=1}^{+\infty}\frac{d_k^n}{n}z^n \tag{3-86}$$

上式中后4项的收敛区域分别为$|z|>a_k$、$|z|>|b_k^{-1}|$、$|z|>c_k$、$|z|>|d_k^{-1}|$。逐项求上式的Z逆变换,可以求得倒谱:

$$\hat{v}(n)=\log|A|\delta(n)-\sum_{k=1}^{m_1}\frac{a_k^n}{n}u(n-1)+\sum_{k=1}^{m_0}\frac{b_k^{-n}}{n}u(-n-1)$$
$$+\sum_{k=1}^{p_1}\frac{c_k^n}{n}u(n-1)-\sum_{k=1}^{p_0}\frac{d_k^{-n}}{n}u(-n-1) \tag{3-87}$$

或写成

$$\hat{v}(n)=\begin{cases}\log|A|, & n=0\\ \sum_{k=1}^{p_1}\frac{c_k^n}{n}-\sum_{k=1}^{m_1}\frac{a_k^n}{n}, & n>0\\ \sum_{k=1}^{m_0}\frac{b_k^{-n}}{n}-\sum_{k=1}^{p_0}\frac{d_k^{-n}}{n}, & n<0\end{cases} \tag{3-88}$$

应该指出,对于有限长度序列,式(3-88)中在n不等于零时的取值将消失。

从上述分析中可以看出,声道响应序列的倒谱特性如下:①倒谱$\hat{v}(n)$是一个双边序列,即在$-\infty<n<\infty$的范围内,$\hat{v}(n)$皆有值;②由于$|a_k|$、$|b_k|$、$|c_k|$、$|d_k|$都小于1,所以倒谱$\hat{v}(n)$是一个衰减序列,即随着$|n|$的增大,$|\hat{v}(n)|$减小,并且衰减速度至少比$\frac{1}{n}$快;③如果信号本身$v(n)$是最小相位序列,即极点和零点皆在Z平面单位圆内部,即$b_k=0$同时$d_k=0$,则$\hat{v}(n)$只在$n\geqslant 0$范围有值,即为因果序列,或者说,最小相位信号序列的倒谱是一个因果序列;④如果$v(n)$是最大相位序列,即极点和零点皆在Z平面单位圆外部,即$a_k=0$同时$c_k=0$,则$\hat{v}(n)$只在$n<0$范围有值,即为反因果序列。或者说,最大相位信号序列的倒谱是一个反因果序列。

实际上,声道的特性取决于式(3-82)的零极点分布。从声道响应的倒谱可知,当$V(z)$的零极点的模值不接近于1时,$\hat{v}(n)$将随着n的增大而迅速递减。当采样频率为10kHz时,$\hat{v}(n)$在间隔[-25,25]之外的值已经相当小,可认为声道响应的倒谱只分布在这一范围内。

3.7.3　避免相位卷绕的算法

在倒谱分析的过程中,由于Z变换后得到的是复数,所以取对数时进行的是复对数的运算。这时将存在相位的多值性问题,形象些说就是将存在“相位卷绕”问题。由于相位卷绕,使得求倒谱及由倒谱恢复语音的运算存在不确定性,因而会产生错误。下面以Z变换是最简单的傅里叶变换运算为例,分析相位卷绕是如何产生的。

设信号

$$x(n) = e(n) * v(n) \tag{3-89}$$

其傅里叶变换为

$$X(\omega) = E(\omega) \times V(\omega) \tag{3-90}$$

复对数如下：

$$\log X(\omega) = \log E(\omega) + \log V(\omega) \tag{3-91}$$

因而有振幅和相位如下：

$$\log | X(\omega) | = \log | E(\omega) | + \log | V(\omega) | \tag{3-92}$$

$$\angle[X(\omega)] = \angle[E(\omega)] + \angle[V(\omega)] \tag{3-93}$$

其中，$\angle$表示求相角。式(3-93)也可以表示为

$$\phi(\omega) = \phi_1(\omega) + \phi_2(\omega) \tag{3-94}$$

式(3-94)表明了相位的多值性，尽管 $\phi_1(\omega)$ 和 $\phi_2(\omega)$单个值是在 0～2π 内。这里由于 $\phi(\omega)$采用了求和，因此其值可能超过 2π。但是，在用计算机计算时，它得到的总相位值$\angle[X(\omega)]$只能用其小于 2π 的主值$\Phi(\omega)$来表示。所以有可能出现

$$\phi(\omega) = \Phi(\omega) + 2\pi k \tag{3-95}$$

其中，k 为整数。由于 k 值无法事先确知，因而真值 $\phi(\omega)$也就无法得出。图 3-25 表示相位卷绕的一个例子。

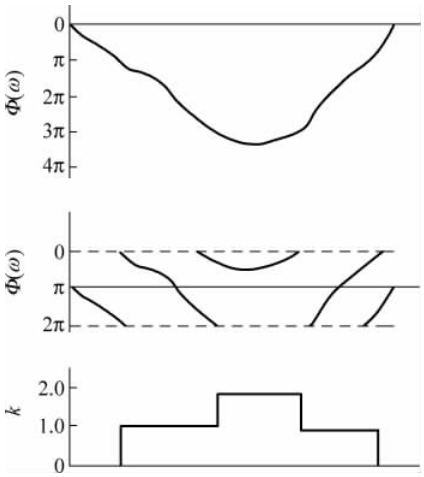

图 3-25　相位卷绕及其校正系数

下面介绍几种避免相位卷绕的方法。

1. 微分法

这种方法利用了傅里叶变换的微分特性和对数微分特性。傅里叶变换的微分特性为

$$\mathrm{j}\frac{\mathrm{d}}{\mathrm{d}\omega}X(\omega) = \sum_{n=-\infty}^{+\infty} nx(n)\mathrm{e}^{-\mathrm{j}\omega n} \tag{3-96}$$

该式表明，若 $x(n)$的傅里叶变换为 $X(\omega)$，则序列 $nx(n)$的傅里叶变换为 $\mathrm{j}\mathrm{d}X(\omega)/\mathrm{d}\omega$。而复倒谱$\hat{x}(n)$和对数谱$\hat{X}(\omega)$之间也满足关系

$$\mathrm{j}\frac{\mathrm{d}}{\mathrm{d}\omega}\hat{X}(\omega) = \sum_{n=-\infty}^{\infty} n\hat{x}(n)\mathrm{e}^{-\mathrm{j}\omega n} \tag{3-97}$$

利用对数微分特性，式(3-97)可以改写为

$$\mathrm{j}\frac{\mathrm{d}}{\mathrm{d}\omega}\hat{X}(\omega) = \mathrm{j}\frac{\mathrm{d}}{\mathrm{d}\omega}[\log X(\omega)] = \mathrm{j}\frac{\frac{\mathrm{d}}{\mathrm{d}\omega}[X(\omega)]}{X(\omega)} = \sum_{n=-\infty}^{+\infty} n\hat{x}(n)\mathrm{e}^{-\mathrm{j}\omega n} \tag{3-98}$$

因此，由式(3-96)和式(3-98)可以画出避免相位卷绕求复倒谱的框图，如图 3-26 所示。

虽然这种方法避免了求复对数的问题，但缺点是会产生严重的混叠。其原因是 $nx(n)$的频谱中的高频分量比 $x(n)$有所增加，所以仍使用 $x(n)$原来的采样率将引起混叠；混叠后求出的$\hat{x}(n)$将不是 $x(n)$的复倒谱。因而这不是一个理想的方法。

2. 最小相位信号法

这是一种较好的解决相位卷绕的方法，它既避开了求复对数过程，又不会产生混叠问

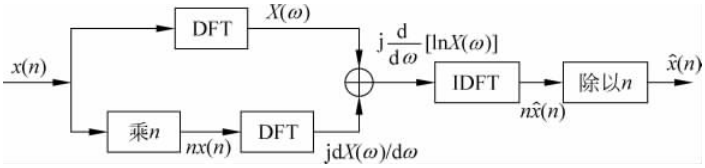

图 3-26　利用微分特性求复倒谱的框图

题。但它有一个限制条件：即被处理的信号 $x(n)$必须是最小相位信号。实际上许多信号都是最小相位信号，或可以看作是最小相位信号。语音信号的模型就是极点都在 Z 平面单位圆内的全极模型，或者极零点都在 Z 平面单位圆内的极零模型。

最小相位信号法是由最小相位信号序列的复倒谱性质，以及希尔伯特(Hilbert)变换的性质推导出来的。设信号 $x(n)$的 Z 变换为 $X(z)=N(z)/D(z)$，则有

$$\hat{X}(z) = \log X(z) = \log \frac{N(z)}{D(z)} \tag{3-99}$$

根据 Z 变换的微分性质有

$$\begin{aligned}\sum_{n=-\infty}^{\infty} n\hat{x}(n)z^{-n} &= -z\frac{\mathrm{d}}{\mathrm{d}z}\hat{X}(z) = -z\frac{\mathrm{d}}{\mathrm{d}z}\left[\log\frac{N(z)}{D(z)}\right] \\ &= \frac{-z\dfrac{\mathrm{d}}{\mathrm{d}z}\left[\dfrac{N(z)}{D(z)}\right]}{\dfrac{N(z)}{D(z)}} = -z\frac{\dfrac{D(z)N'(z)-N(z)D'(z)}{D^2(z)}}{\dfrac{N(z)}{D(z)}} \\ &= -z\frac{D(z)N'(z)-N(z)D'(z)}{N(z)D(z)}\end{aligned} \tag{3-100}$$

如果 $x(n)$是最小相位信号，则 $N(z)$和 $D(z)$的所有根均在 Z 平面的单位圆，$n\hat{x}(n)$的 Z 变换的所有极点也均位于 Z 平面单位圆内。这表明，若 $x(n)$是最小相位信号，则$\hat{x}(n)$必然是稳定的因果序列。

另一方面，由希尔伯特变换的性质可知，任一因果的复倒谱序列$\hat{x}(n)$都可以分解为偶对数分量$\hat{x}_{\mathrm{e}}(n)$和奇对数分量$\hat{x}_{\mathrm{o}}(n)$之和，即

$$\hat{x}(n) = \hat{x}_{\mathrm{e}}(n) + \hat{x}_{\mathrm{o}}(n) \tag{3-101}$$

而且，这两个分量的傅里叶变换分别为$\hat{x}(n)$的傅里叶变换的实部和虚部。设

$$\hat{X}(\omega) = \sum_{n=-\infty}^{+\infty} \hat{x}(n)\mathrm{e}^{-\mathrm{j}\omega n} = \hat{X}_R(\omega) + \mathrm{j}\hat{X}_I(\omega) \tag{3-102}$$

则

$$\hat{X}_R(\omega) = \sum_{n=-\infty}^{+\infty} \hat{x}_{\mathrm{e}}(n)\mathrm{e}^{-\mathrm{j}\omega n} \tag{3-103}$$

$$\hat{X}_I(\omega) = \sum_{n=-\infty}^{+\infty} \hat{x}_{\mathrm{o}}(n)\mathrm{e}^{-\mathrm{j}\omega n} \tag{3-104}$$

图 3-27 给出了将复倒谱因果序列$\hat{x}(n)$分解为$\hat{x}_{\mathrm{e}}(n)$和$\hat{x}_{\mathrm{o}}(n)$的情况。由图可见，它们可由$\hat{x}(n)$和$\hat{x}(-n)$求得

$$\hat{x}_{\mathrm{e}}(n) = \frac{1}{2}[\hat{x}(n) + \hat{x}(-n)] \tag{3-105}$$

$$\hat{x}_o(n) = \frac{1}{2}[\hat{x}(n) - \hat{x}(-n)] \tag{3-106}$$

由此可得

$$\hat{x}(n) = \begin{cases} 0, & n < 0 \\ \hat{x}_e(n), & n = 0 \\ 2\,\hat{x}_e(n), & n > 0 \end{cases} \tag{3-107}$$

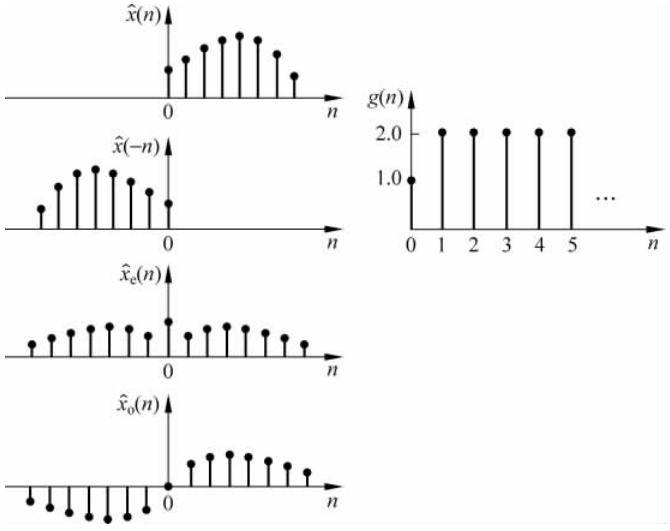

图 3-27 因果序列的分解和恢复

这表明,一个因果序列可由其偶对称分量来恢复。如果引入一个辅助因子 $g(n)$,则上式可以写为

$$\hat{x}(n) = g(n)\,\hat{x}_e(n) \tag{3-108}$$

式中

$$g(n) = \begin{cases} 0, & n < 0 \\ 1, & n = 0 \\ 2, & n > 0 \end{cases} \tag{3-109}$$

根据上述原理,可以画出最小相位法求复倒谱的原理框图,如图 3-28 所示。

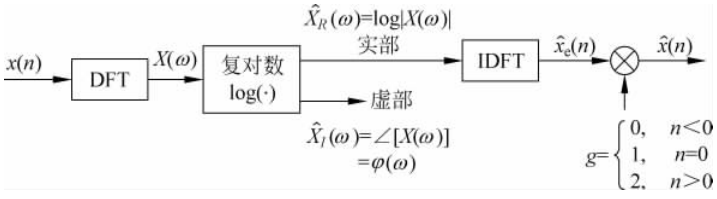

图 3-28 最小相位信号法求复倒谱

3. 递归法

这也是一种避开相位卷绕而能从 $x(n)$求出$\hat{x}(n)$的方法。它也仅限于 $x(n)$是最小相位

信号的情况。所谓递归是指在运算$\hat{x}(n)$时，除了要已知$x(n)$之外，还要知道在$n'<n$时$\hat{x}(n')$各值。根据Z变换的微分特性，有

$$-z\frac{\mathrm{d}}{\mathrm{d}z}\hat{X}(z)=-z\frac{\mathrm{d}}{\mathrm{d}z}[\log X(z)]=-z\frac{\frac{\mathrm{d}}{\mathrm{d}z}X(z)}{X(z)} \tag{3-110}$$

得

$$-zX(z)\frac{\mathrm{d}}{\mathrm{d}z}\hat{X}(z)=-z\frac{\mathrm{d}}{\mathrm{d}z}X(z) \tag{3-111}$$

对上式求Z逆变换，根据Z变换的微分性质，有

$$[n\hat{x}(n)]*x(n)=nx(n) \tag{3-112}$$

或写为

$$\sum_{k=-\infty}^{\infty}[k\hat{x}(k)]x(n-k)=nx(n) \tag{3-113}$$

所以

$$x(n)=\sum_{k=-\infty}^{\infty}\left(\frac{k}{n}\right)\hat{x}(k)x(n-k),\quad n\neq 0 \tag{3-114}$$

设$x(n)$是最小相位信号序列，而最小相位信号序列一定为因果序列，所以有

$$\begin{cases}x(n)=0, & n<0\\ \hat{x}(n)=0, & n<0\end{cases} \tag{3-115}$$

此时可以将$x(n)$写作

$$x(n)=\sum_{k=0}^{n}\left(\frac{k}{n}\right)\hat{x}(k)x(n-k)=\sum_{k=0}^{n-1}\left(\frac{k}{n}\right)\hat{x}(k)x(n-k)+\hat{x}(n)x(0) \tag{3-116}$$

其中，由于当$k<0$时，$\hat{x}(k)=0$；且在$k>n$时$x(n-k)=0$，所以求和的上下限变为由0到n。由此得到的递归公式为

$$\hat{x}(n)=\frac{x(n)}{x(0)}-\sum_{k=0}^{n-1}\left(\frac{k}{n}\right)\hat{x}(k)\frac{x(n-k)}{x(0)},\quad n>0 \tag{3-117}$$

在实际应用中，一般只知道$x(n)$，并不知道在$n'<n$时$\hat{x}(n')$。但是可以在第一次递归之前先求出$\hat{x}(0)$，这样就可以进行递归运算。求$\hat{x}(0)$的方法如下，由复倒谱定义

$$\hat{x}(n)=Z^{-1}\{\log Z[x(n)]\}=Z^{-1}\left\{\log\left[\sum_{n=-\infty}^{\infty}x(n)z^{-n}\right]\right\} \tag{3-118}$$

在$n=0$时

$$\hat{x}(0)=Z^{-1}[\log x(0)]=\log x(0)\delta(n)\big|_{n=0}=\log x(0) \tag{3-119}$$

顺便指出，如果$x(n)$是最大相位序列，则式(3-109)中的$g(n)$为

$$g(n)=\begin{cases}0, & n>0\\ 1, & n=0\\ 2, & n<0\end{cases} \tag{3-120}$$

而这时递归公式变成

$$\hat{x}(n)=\frac{x(n)}{x(0)}-\sum_{k=n+1}^{0}\left(\frac{k}{n}\right)\hat{x}(k)\frac{x(n-k)}{x(0)},\quad n<0 \tag{3-121}$$

3.7.4　基于听觉特性的Mel频率倒谱系数

在语音识别和说话人识别中，常用的语音特征是基于Mel频率的倒谱系数(mel

frequency cepstrum coefficient,MFCC)。由于 MFCC 参数是将人耳的听觉感知特性和语音的产生机制相结合,因此大多数语音识别系统中广泛使用这种特征。

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

$$f_{\text{Mel}} = 2595 \times \lg(1 + f/700) \tag{3-122}$$

对频率轴的不均匀划分是 MFCC 特征区别于前面所述的普通倒谱特征的最重要的特点。将频率按照式(3-122)变换到 Mel 域后,Mel 带通滤波器组的中心频率是按照 Mel 频率刻度均匀排列的。在实际应用中,MFCC 倒谱系数计算过程如下:

(1) 将信号进行分帧,预加重和加汉明窗处理,然后进行短时傅里叶变换得到其频谱;

(2) 求出频谱平方,即能量谱,并用 M 个 Mel 带通滤波器进行滤波,由于每一个频带中分量的作用在人耳中是叠加的,因此将每个滤波频带内的能量进行叠加,这时第 k 个滤波器输出功率谱 $x'(k)$;

(3) 将每个滤波器的输出取对数,得到相应频带的对数功率谱;并进行反离散余弦变换,得到 L 个 MFCC 系数,一般 L 取 12～16,如下式所示:

$$C_n = \sum_{k=1}^{M} \log x'(k) \cos[\pi(k-0.5)n/M], \quad n = 1,2,\cdots,L \tag{3-123}$$

(4) 这种直接得到的 MFCC 特征作为静态特征,将这种静态特征做一阶和二阶差分,得到相应的动态特征。

表 3-3 给出了 13 维 MFCC 特征及其动态特征对系统识别性能的影响。

表 3-3 动态特征对系统识别性能的影响

特征集合	相对误识率的降低	特征集合	相对误识率的降低
13 维的 LPCC 特征	基线系统	1 阶和 2 阶动态特征	+20%
13 维的 MFCC 特征	+10%	3 阶动态特征	+0%
16 维的 MFCC 特征	+0%		

表 3-3 以 13 维的 LPCC 倒谱特征为基线系统,可以看出,MFCC 系统由于有效利用了听觉特性,因此其改进了识别系统性能。如果将倒谱维数增加,对识别性能影响不大,误识率基本上与 13 维时一样。但采用动态特征,误识率可以有 20%的下降。动态阶数继续增加时,其性能没有进一步提高。

3.8 语音信号特征应用

前面各节介绍了语音信号的时域特征、频域特征,以及一些可直接用于语音信号处理的其他特征等。此外,语音信号中还有一些如共振峰和基音周期等固有特征,本节将对这些问题加以介绍。

3.8.1　基音周期估计

基音是指发浊音时声带振动所引起的周期性，而基音周期是指声带振动频率的倒数。由于它只是准周期性的，所以只能采用短时平均方法估计其周期，这个过程也常称为基音检测(pitch detection)。

基音周期是语音信号最重要的参数之一，它的提取是语音信号处理中一个十分重要的问题，尤其是对汉语更是如此；因为汉语是一种有调语言，基音的变化模式称为声调。声调携带着非常重要的具有辨意作用的信息，有区别意义的功能。根据加窗的短时语音帧来估计基音周期，在语音编解码器、语音识别、说话人确认和辨认，以及生理缺陷人的辅助系统等许多领域都是重要的一环。自进行语音信号分析研究以来，基音检测一直是一个重点研究的课题，已经提出了很多方法，然而这些方法都有它们的局限性。迄今为止，尚未找到一个完善的可以适用于不同的说话人、不同的要求和环境的基音检测方法。

基音检测的主要困难表现在：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周期的序列，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是很难判定它应属于周期性或非周期性，从而也就无法估计出基音周期；②要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，例如声道共振峰有时会严重影响激励信号的谐波结构；③在浊音段很难精确地确定每个基音周期的开始和结束位置，这不仅因为语音信号本身是准周期的，也是因为波形的峰受共振峰结构、噪声等影响；④基音周期变化范围较大，从低音(男声)80Hz 直到(女孩)500Hz，也给基音周期的检测带来了一定的困难。另外，浊音信号可能包含有 30～40 次谐波分量，而基波分量往往不是最强的分量。因为语音的第一共振峰通常在 300～1000Hz 范围内，这就是说，2～8 次谐波成分往往比基波分量还强。丰富的谐波成分使语音信号的波形变得很复杂，给基音检测带来困难，经常发生基频估计结果为实际基音频率的二、三次倍频或二次分频的情况。

基音检测的方法大致可分为三类：①波形估计法，直接由语音波形来估计基音周期，分析出波形上的周期峰值，包括并行处理法、数据减少法等；②相关处理法，这种方法在语音信号处理中广泛使用，这是因为相关处理法抗波形的相位失真能力强，另外它在硬件处理上结构简单，包括波形自相关法、平均振幅差分函数法(AMDF)、简化逆滤波法(SIFT)等；③变换法，将语音信号变换到频域或倒谱域来估计基音周期，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，进一步求取基音周期，比如倒谱法。虽然倒谱分析算法比较复杂，但基音估计效果较好。各种方法的对比见表 3-4 所示。

表 3-4　典型的基音周期检测方法

分　类	基音检测方法	特　　点
波形估计法	并行处理方法	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
	过零率法	关于波形的过零率，着眼于重复图形

续表

分　类	基音检测方法	特　　点
相关处理法	自相关法及其改进	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 计算法	语音波形降低采样后，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	采用平均幅差函数检测周期性，也可以根据残差信号的 AMDF 进行提取
变换法	倒谱法	根据对数功率谱的傅里叶反变换，分离频谱包络和微细结构
	循环直方图	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

下面介绍常用的几种基音检测方法。

1. 自相关方法

浊音信号的自相关函数在基音周期的整数倍位置上出现峰值，而清音的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或浊音，峰-峰值之间对应的就是基音周期。

影响从自相关函数中正确提取基音周期的最主要原因是声道响应部分。当基音的周期性和共振峰的周期性混在一起时，被检测出来的峰值就可能会偏离原来峰值的真实位置。另外，某些浊音中，第一共振峰频率可能会等于或低于基音频率。此时，如果其幅度很高，它就可能在自相关函数中产生一个峰值，而该峰值又可以同基音频率的峰值相比拟。

为了提高自相关方法检测基音周期的准确性，需要进行一些前期的预处理。

1）预处理

语音信号的低幅值部分包含大量的共振峰信息，而高幅值部分包含较多的基音信息。因此，任何削减或者抑制语音低幅度部分的非线性处理都会使自相关方法的性能得到改善。中心削波即是一种非线性处理，它消除语音信号的低幅度部分，其削波特性如图 3-29 所示，数学表达形式为

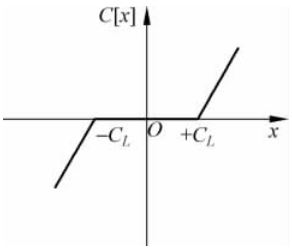

图 3-29　中心削波函数

$$y(n)=C(n)=\begin{cases}x(n)-L, & x(n)>C_L\\ 0, & |x(n)|\leqslant C_L\\ x(n)+L, & x(n)<-C_L\end{cases} \tag{3-124}$$

式中，削波电平 C_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的 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，以达到较好的效果。经过中心削波后只保留了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，而其余的次要峰值幅度都很小。据报道使用这种方法，对电话带宽的语音在信噪比低至 18dB 的情况下获得了良好的性能。

计算自相关函数的运算量是很大的，其原因是传统的计算机进行乘法运算非常费时。尽管近年来随着数字信号处理器的广泛使用，实时地计算自相关函数已经不是问题，但在基音检测中仍然有一些减少短时自相关运算的有效方法。例如可对中心削波函数进行修正，

采用三电平中心削波的方法，如图 3-30 所示。其削波函数为

$$y(n)=C[x(n)]=\begin{cases}1, & x(n)>C_L\\ 0, & |x(n)|\leqslant C_L\\ -1, & x(n)<-C_L\end{cases}\quad(3\text{-}125)$$

即削波器的输出在 $x(n)>C_L$ 时为 1，$x(n)<-C_L$ 时为 -1，除此以外均为零。虽然这一处理会增加刚刚超过削波电平峰的重要性，但大多数次要的峰被滤除掉了，而只保留了明显的周期性峰。

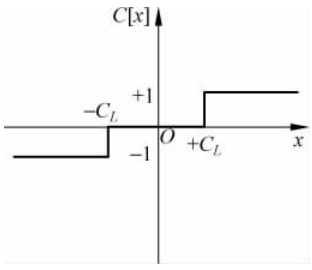

图 3-30　三电平削波函数

此外，还可以用一个通带为 900Hz 的线性相位低通滤波器滤除高次谐波分量。这样处理后的信号，基本上只含有第一共振峰以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法的基音检测都有明显的效果。

2）基于自相关函数的基音检测

短时自相关函数在基音周期的各个整数倍点上有很大的峰值，只要找到第一最大峰值点的位置，并计算它与原点的间隔，便能估计出基音周期。但实际上并不是这么简单，第一个最大峰值点的位置有时不能与基音周期相吻合。产生这种情况的原因有两个方面：一方面与窗的长度有关，一般认为窗长至少应大于两个基音周期，才可能有较好的效果；另一方面与声道特性的影响有关，有的情况下，即使窗长已经选得足够长，第一个最大峰值点与基音周期仍不一致，这就是声道共振峰特性的干扰。经过上述带通滤波的预处理，可以消除大部分的共振峰的影响。但是，如果希望减少自相关计算中的乘法运算，可以把上述中心削波后的信号$\{y(n)\}$的自相关序列用两个信号的互相关序列代替，其中一个信号是$\{y(n)\}$，另一个信号是对$\{y(n)\}$进行三电平量化产生的结果$\{y'(n)\}$。显然，$y'(n)$只有$-1,0,+1$三种可能的取值，因而这里的互相关计算只需做加减法，而这个互相关序列的周期性与$\{y(n)\}$的自相关序列近似相同。

下面结合 L. R. Rabiner 在一篇论文中介绍的具体例子来叙述关于自相关函数的基音检测方法。假设信号的采样率为 10kHz，窗序列采用 300 点的矩形窗，帧叠 200 点。这时对每一帧进行基音周期估计的步骤如下：

（1）用 900Hz 低通滤波器对一帧语音信号$\{x(n)\}$进行滤波，并去掉开头的 20 个输出值不用，得到$\{x'(n)\}$。

（2）分别求$\{x'(n)\}$的前部 100 个样点和后部 100 个样点的最大幅度，并取其中较小的一个，乘以因子 0.68 作为门限电平 C_L。

（3）对$\{x'(n)\}$分别进行中心削波得到$\{y(n)\}$和三电平量化得到$\{y'(n)\}$。

（4）求这两个信号的互相关值 $R(k)$。其中 $R(k)=\sum_{n=21}^{300}y(n)\cdot y'(n+k)$，此处 k 的取值范围 20～150 相应于基音频率范围 60～500Hz，$R(0)$相应于短时能量。

（5）得到互相关值后，可以得到 $R(20)\cdots R(150)$中的最大值 $R_{\max}$，如果 $R_{\max}<0.25R(0)$，则认为本帧为清音，令其基音周期值为 0，否则基音周期即为使 $R(k)$为最大值 $R_{\max}$时位置 k 的值，即 $p=\underset{20\leqslant k\leqslant 150}{\operatorname{argmax}}R(k)$。

2. 基于短时平均幅度差的基音周期估计

平均幅度差函数只涉及加减和求绝对值运算，因此不需要做中心削波和三电平量化。

首先，只要将一帧信号$\{x(n)\}$经过900Hz低通滤波器处理后得到$\{x'(n)\}$；计算$\{x'(n)\}$的平均幅度差函数$\gamma(k)$，并求出取得这一最小值时的下标作为基音周期的初步值，即$p=\underset{k}{\operatorname{argmin}}\gamma(k)$。这时的平均幅度差函数的最小值为$\gamma_{\min}=\underset{k}{\min}\gamma(k)$。其次，搜寻平均幅度差函数的若干局部极小值点作为基音周期的候选。这些局部极小值点必须满足两个条件：①其取值应在$\gamma_{\min}\sim\gamma_{\min}+\gamma_{TH}$的范围内，$\gamma_{TH}$是一个恰当选取的阈值；②各个局部极小值点之间的间隔不得小于l_{TH}，l_{TH}是一个恰当选取的间隔值，在实际应用中要根据实验确定。对于各个局部极小值点进行再度检查，确定清晰点。在某个最小点左右各8个点范围内对平均幅度差函数求平均，若该最小点与此平均值的差距大于某个阈值γ_D，称为清晰点；最后，在所有清晰点中找到最左边的那个点，就是该帧语音的基音周期值。

3. 倒谱法

对语音信号利用倒谱解卷原理，可以得出激励序列的倒谱，它具有与基音周期相同的周期，因此可以容易且精确地求出基音周期。图3-31(a)为语音信号对数频谱示意图，它包含两个分量：对应于频谱包络的慢变分量(如虚线所示)，以及对应于基音谐波峰值的快变分量(如实线所示)。通过滤波或再取一次傅里叶逆变换，即可将慢变分量与快变分量分离开。图3-31(b)为倒谱$c(n)$的示意图，其中靠近原点的低倒频部分是频谱包络的变换，而位于t_0处的窄峰为谐波峰值的变换，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，从而能对前者很容易地加以识别。

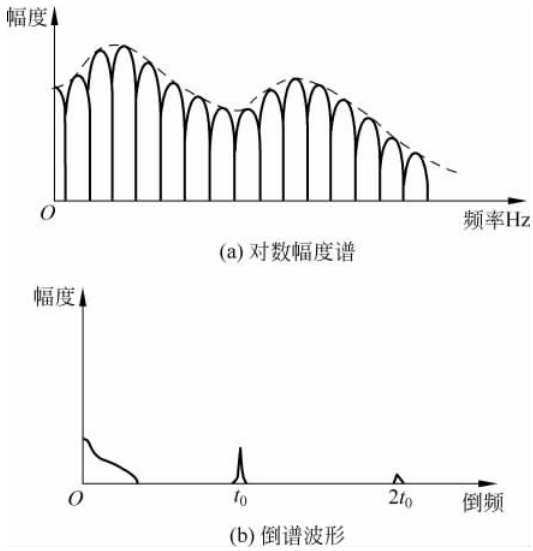

图3-31 倒谱示意图

下面举一个用倒谱提取基音的实例，如图3-32所示，其工作原理简要说明如下。

(1) 采样率为10kHz，帧长51.2ms，用汉明窗平滑，然后求出倒谱。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，窗口选择的语音段应至少包含有两个明显的周期。例如对基音频率低的男性，要求窗口长度为40ms；而对基音频率高的语音，窗的长度可以成比例地缩短。

(2) 求出倒谱峰值I_{pk}及其位置I_{pos}，如果峰值未超过某门限值，则进行过零计算；若过

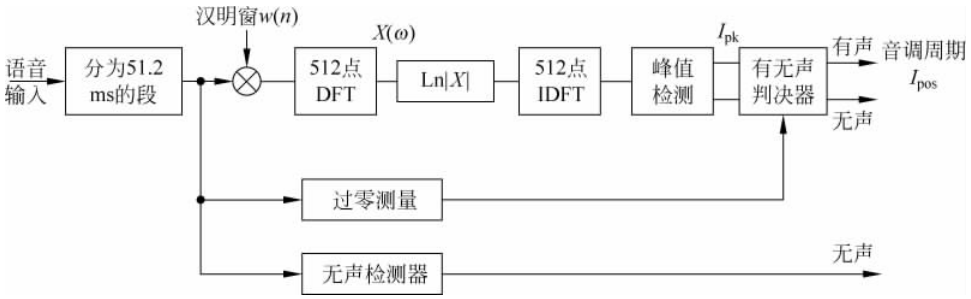

图 3-32　基音检测的倒谱法

零率低于某门限值，则为无声语音帧。反之，则为有声语音帧，且基音周期仍等于该峰值的位置。

(3) 图中的无声检测器是时域信号的峰值检测器；若低于某门限值，则认为是无声，不进行上述由倒谱检测基音的计算。

当采用无噪语音时，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展宽，并受到噪声的污染，从而使倒谱的灵敏度也随之下降。

4. 简化逆滤波法

简化的逆滤波跟踪算法先抽取声道模型参数，利用这些参数对原信号进行逆滤波，从预测误差中得到声源序列，再用自相关法求得基音周期。语音信号通过线性预测逆滤波器后达到频谱的平坦化。预测误差是自相关器的输入，通过与门限的比较可以确定浊音，通过辅助信息可以减少误差。

简化逆滤波器的原理框图如图 3-33 所示，其工作过程如下：

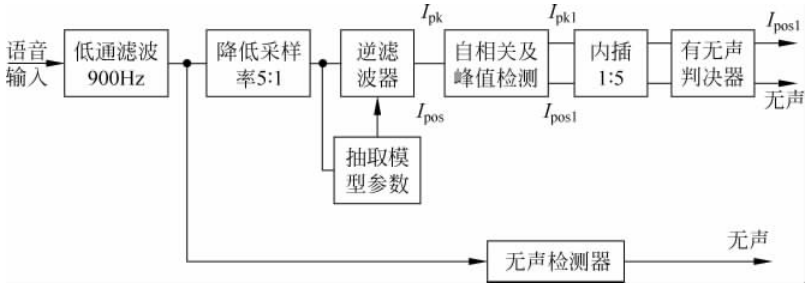

图 3-33　简化逆滤波法原理

(1) 语音信号经过 10kHz 采样后，通过 0～900Hz 的低通滤波器，然后将采样率降低为原采样率的 1/5(因为激励序列的宽度小于 1kHz，所以用 2kHz 采样就足够了)；当然，后面要进行内插。

(2) 提取降低采样率后的信号模型参数(LPC 参数，见第 4 章)，利用声道模型参数构造一个逆滤波器。经过逆滤波器后的信号是与声道特性分离的激励源信号，经过相应的自相关算法后，检测出峰值及其位置，就得到基音周期值。

(3) 最后进行有/无声判决。与前面倒谱法类似,有一个无声检测器,以减少运算量。

在基音检测中,广泛采用对语音波形或误差信号波形进行低通滤波,因为这种低通滤波对提高基音检测精度有良好的效果。低通滤波在去除了高阶共振峰影响的同时,还可以补充自相关函数时间分辨率的不足。特别是后者在用线性预测误差的自相关函数的基音检测中尤其重要。

无论采用哪一种算法求得的基音周期轨迹与真实的基音周期轨迹不可能完全一致。实际情况是大部分段落是一致的,而在一些局部段落或区域中有一个或几个基音周期的估计值偏离了正常的轨迹(通常是偏离到正常值的 2 倍或 1/2),这时称为基音轨迹产生了若干"野点"。为了去除这些野点,可以采用各种平滑算法,其中最常用的是中值平滑算法和线性平滑算法。

在中值滤波平滑算法中,被平滑点的左右各取 L 个样点,连同被平滑点共同构成一组 $2L+1$ 个信号样点值。将这些样点值按大小次序排成一队,取此队列中间者作为平滑器的输出。L 值一般取为 1 或 2,即中值平滑的"窗口"一般套住 3 或 5 个样值。中值平滑的优点是既可以有效地去除少量野点,又不会破坏基音周期轨迹中的两个平滑段之间的阶跃性变化。

线性平滑是用滑动窗进行线性滤波处理,即

$$y(n)=\sum_{m=-L}^{L}x(n-m)\cdot w(m) \tag{3-126}$$

式中,$\{w(m),m=-L,-L+1,\cdots,0,\cdots,L\}$为 $2L+1$ 点平滑窗,满足

$$\sum_{m=-L}^{L}w(m)=1 \tag{3-127}$$

例如,三点窗的值可取为$\{0.25,0.5,0.25\}$。线性平滑在纠正输入信号中不平滑处样点的同时,也使附近的样点值做了修改,所以窗长不易过大。

3.8.2 共振峰的估计

共振峰是反映声道谐振特性的重要特征,它代表了发音信息的最直接的来源,而且人在语音感知中也利用了共振峰信息。所以共振峰是语音信号处理中非常重要的特征参数。

共振峰信息包含在语音频谱包络中,因此提取共振峰参数的关键是估计语音的频谱包络,一般认为谱包络中的最大值就是共振峰。与基音检测类似,共振峰估计也是表面上看起来很容易,而实际上又受许多问题困扰。这些问题包括以下几类。

(1) 虚假峰值。在正常情况下,频谱包络中的极大值完全是由共振峰引起的。但在线性预测分析方法出现之前的频谱包络估计器中,出现虚假峰值是相当普遍的现象。甚至在采用线性预测方法时,也并非没有虚假峰值。为了增加灵活性会给预测器增加 2～3 个额外的极点,有时可利用这些极点代表虚假峰值。

(2) 共振峰合并。相邻共振峰的频率可能会靠得太近而难以分辨。这时会产生共振峰合并现象,而探讨一种理想的能对共振峰合并进行识别的共振峰提取算法存在很多实际困难。

(3) 高音调语音。传统的频谱包络估计方法是利用由谐波峰值提供的样点。高音调语音(如女声和童声)的谐波间隔比较宽,因而为频谱包络估值所提供的样点比较少,所以谱包

络本身的估计就不够精确。即使采用线性预测进行频谱包络估计也会出现这个问题。在这样的语音中，线性预测包络峰值趋向于离开真实位置，而朝着最接近的谐波峰位移动。

下面讨论常用的几种共振峰提取方法。

1. 基于线性预测的共振峰求取方法

一种有效的频谱包络估计方法是从线性预测分析角度推导出声道滤波器，根据这个声道滤波器找出共振峰。虽然线性预测法也有一定的缺点，例如其频率灵敏度与人耳不相匹配，但对于许多应用来说，它仍然是一种行之有效的方法。线性预测共振峰估计通常有两种途径可供选择：一种途径是利用一种标准的寻找复根的程序计算预测误差滤波器的根，称为求根法；另一种途径是找出由预测器导出的频谱包络中的局部极大值，称为选峰法。

1）求根法

这种方法是找出多项式复根，根据求得的根来确定共振峰。通常采用牛顿-拉夫逊(Newton-Raphson)搜索算法。该算法一开始先猜测一个根值，并就此猜测值计算多项式及其导数的值，然后利用计算结果再找出一个改进的猜测值。通常当前后两个猜测值之差小于某个事先设定的阈值时，结束求根过程。

若求出的根为实根，则在多项式中相对应的因子项是线性的；若为复根，则通过该根及其共轭可以找到一个二次因子。通过使多项式降阶有效地去掉这个根，然后利用上面的求根方法，求出降阶后多项式的与此不同的根。多项式降阶与求根过程如此重复进行下去，直到将全部的根找出为止。由于被去掉的根并不是精确已知的，从而导致多项式降阶总要造成某些精度的损失，因而用这种方法相继求出的根在精度方面越来越差。避免这个问题的方法通常是对于未降阶多项式的每一个新根实行最后的牛顿-拉夫逊重复运算。有时利用这个算法可能会找到远离单位圆的猜测值，这时可以将猜测值到原点的距离限制在某个合适的范围之内。对于自相关预测器，极点总是位于单位圆内；而对于协方差预测器，即使在最坏的情况下，极点也只是在一个短距离之外，因此上述限制并不妨碍从已找出的根得到修正根。

假如每一帧的最初猜测值与前一帧的根的位置重合，那么一般来说根的帧到帧的移动足够小，经过较少的重复运算之后，即可使新的根值会聚在一起。当求根过程刚开始的时候，第一帧的最初猜测值可以在单位圆上等间隔放置。

如果在某个点 z_i 是一个根，那么与 i 对应的共振峰频率和三分贝带宽分别由下面公式给出：

$$F_i = \frac{\theta_i}{2\pi T_s} \tag{3-128}$$

$$B_i = \frac{\ln |z_i|}{\pi T_s} \tag{3-129}$$

其中，$T_s = 1/f_s$。例如，若求出一个根位于 $z_i = 0.1 + j0.95$，则 $|z_i| = 0.955$，$\theta_i = 1.466$。若语音的采样频率为 8kHz，则共振峰频率为 $F_i = 1866$Hz，三分贝带宽 $B_i = 117$Hz。因为极点是以共轭对形式出现的，所以只需要对虚部为正的极点进行考察就可以。若 B_i 为负值，则相应的极点位于单位圆外。这时对 B_i 的修正，通常可以用 $1/z_i$ 代替 z_i，即可将极点反射到单位圆内，显然这样做并不影响 B_i 的绝对值。

对于实时语音处理来说，多项式求根的计算开销通常是很大的，一般不可取。但这种方

法可以用于实验研究。

2）选峰法

由预测器系数获得共振峰数据的另一个途径是计算出语音信号的频谱包络，然后通过对频谱包络中局部极大值进行搜索找出共振峰。显然选峰法比求根法容易实现。选峰法的主要缺点是对共振峰合并现象无能为力，对于共振峰合并来说，两个相邻共振峰的极点紧紧地靠在一起，从而频谱包络只呈现出一个局部极大值，而不是两个极大值。于是峰值检测器认为在此处只存在一个共振峰，当将峰值同共振峰对号入座时便会引起一系列的混乱。

解决共振峰合并问题最有效的方法是减少从极点到计算频谱包络曲线的距离。显然，如果极点位于单位圆内，并通过在单位圆与极点之间的曲线上对函数求值，那么所得到的频谱包络也就不大可能出现共振峰合并。原则上说，只要用于函数求值的曲线和极点相距足够近，那么任何共振峰合并问题都可以解决。

利用频谱包络中局部极大值进行搜索寻找共振峰，会将谐波峰值误识为共振峰。下面介绍一种利用谐波频率及其上下两个次极值频率求得共振峰频率的方法。

设激励频率为 F_0，则语音信号的频谱将出现多个谐波频率 $f=nF_0$，它们的位置是频谱曲线的各峰值处。图 3-34 表示如何从谐波频率求得共振峰频率的两种内插关系，即可由谐波频率 f 及其上下两个次极值频率 $f+F_0$、$f-F_0$ 的插值来求得共振峰频率：

$$F = f \pm \Delta f \tag{3-130}$$

其中，Δf 是谐波频率与共振峰频率之差。

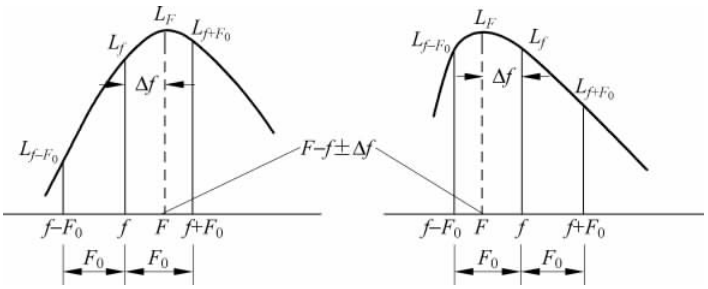

图 3-34　共振峰频率与谐波频率之间的关系

具体内插时的几何关系如图 3-35 所示。

由图 3-35，可知

$$\frac{d_2}{BG} = \frac{d_1}{F_0}, \quad BG + (F_0 - \Delta f) = F_0 + \Delta f \tag{3-131}$$

因此有

$$\Delta f = \frac{d_2}{2d_1} F_0 \tag{3-132}$$

即可以得到两种内插可能的共振峰频率：

$$F = f \pm \frac{d_2}{2d_1} F_0 \tag{3-133}$$

共振峰幅值 L_F 与谐波频率时幅值 L_f 之差是 ΔL，则由图 3-35 的几何关系及式(3-133)可以得到 $d_1/F_0=\Delta L/\Delta f$，因此有

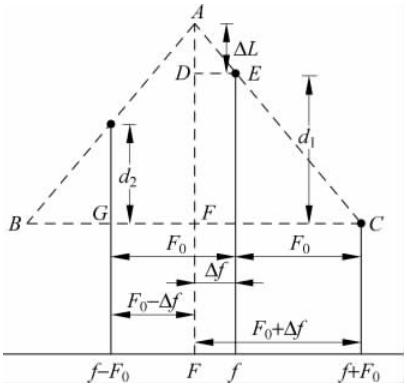

图 3-35　计算共振峰频率的图解

$$\Delta L = \frac{d_2}{2}, \quad L_F = L_f + \frac{d_2}{2} \tag{3-134}$$

而共振峰带宽如图 3-35 中的 d_1 用分贝表示，由于$\left(\frac{BF}{2}\right)/3=\frac{F_0}{d_1}$，故共振峰三分贝带宽可以表示为

$$B_F = 6F_0/d_1 \tag{3-135}$$

2. 倒谱法

从前面的同态分析可以知道，由于声道响应的倒谱衰减很快，在[－25，25]之外的值已经相当小，因此可以构造一个相应的倒谱滤波器，将声道的倒谱分离。对分离出来的倒谱做相应的反变换，就可以得到声道函数的对数谱，对此做进一步处理即可求得所需的各个共振峰。需要注意，实际分析中的语音信号是一段加窗的短时语音。未加窗的语音信号 $x(n)$等于激励信号 $e(n)$和声道响应 $v(n)$的卷积。而加窗信号 $x_w(n)$可以表示为 $x_w(n)=[e(n)*v(n)]w(n)$，式中 $w(n)$为某种窗函数。可以从频域或时域角度估计加窗对同态分析的影响。

由于 $x_w(n)$等于 $x(n)$和 $w(n)$的乘积，$x_w(n)$的频谱等于 $x(n)$的频谱与 $w(n)$的频谱的卷积，由此引入的畸变主要来自 $w(n)$频谱的主瓣宽度不够窄和主瓣以外的波纹造成的泄漏现象。为了克服后者，窗函数一般选为汉明窗，而很少用方窗。对于前者，当语音帧的长度为 20ms 左右时，所引入的畸变不是很大，因此可以接受。

从时域角度，$x_w(n)$可以写成

$$x_w(n) = \left[\sum_{l=-\infty}^{+\infty} v(n)e(n-l)\right]w(n) \tag{3-136}$$

考虑到 $v(n)$是声道函数的单位取样响应，是因果序列，所以对持续时间也有限制。因此 $v(n)$的非零间隔可以表示为$[0,n_l]$，式中 n_l 是一个与语音短时帧的点数相比小得多的正整数。再假设 $w(n)$的变化在$[0,n_l]$范围内，因此当 $l\in[0,n_l]$时 $w(n_l)\approx w(n)$。这样，语音信号可用下面公式近似表示。

$$x_w(n)=\left[\sum_{l=0}^{n_l} v(l)e(n-l)\right]w(n) = \sum_{l=0}^{n_l}\{v(l)\}\{e(n-l)w(n)\}$$

$$\approx \sum_{l=0}^{n_l} \{v(l)\}\{e(n-l)w(n-l)\} \tag{3-137}$$

设 $e_w(n)=e(n)w(n)$，就可以得到

$$x_w(n) \approx \sum_{l=0}^{n_l} v(l)e_w(n-l) = v(n) * e_w(n) \tag{3-138}$$

这样，对加窗语音进行同态分析，并采用倒谱滤波器分离，就可以得到 $v(n)$ 和 $e_w(n)$。从而可以由此确定共振峰及其声道和激励参数。在此讨论中所做的重要假设是 $w(n)$ 必须变化比较缓慢。汉明窗的变化缓慢，而方窗的变化剧烈，从这一角度出发也应该选择前者，这与在频域的讨论结果一致。

参考文献

[1] 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.

[2] 陈永彬，王仁华. 语言信号处理[M]. 合肥：中国科技大学出版社，1990.

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

[4] 冉启文. 小波变换与分数傅里叶变换理论及应用[M]. 哈尔滨：哈尔滨工业大学出版社，2001.

[5] 刘贵忠，邸双亮. 小波分析及其应用[M]. 西安：西安电子科技大学出版社，1995.

[6] 程正兴. 小波分析算法与应用[M]. 西安：西安交通大学出版社，1998.

[7] 王宏禹. 非平稳随机信号分析与处理[M]. 北京：国防工业出版社，1999.

[8] 王炜，杨道淳，方元，等. 基于听觉模型的小波包变换的语音增强[J]. 南京大学学报(自然科学)，2001，37(5)：630-636.

[9] 陶传会，杨道淳，王炜. 听觉系统识别语音信号的模拟[J]. 数据采集与处理，1999，14(2)：157-162.

[10] 陈东义，曹长修，朱冰莲，等. 一种简化的小波去噪算法[J]. 重庆大学学报(自然科学版)，1997，20(5)：63-67.

[11] 张维强. 小波分析及其在语音信号处理中的应用[D]. 西安：西安电子科技大学，2000.

[12] 胡惠英，吴善培. 小波去噪在语音识别中的应用[J]. 北京：北京邮电大学学报，1999，22(3)：31-34.

[13] 陈尚勤，罗成烈，杨雪. 近代语音识别[M]. 成都：电子科技大学出版社，1991.

[14] 胡光锐. 语音处理与识别[M]. 上海：上海科学技术文献出版社，1994.

[15] Xuedong Huang，Alex Acero，Hsiao-Wuen Hon. Spoken Language Processing：A Guide to Theory，Algorithm and System Development[M]. New Jersey：Prentice Hall，2001.