视听觉信号处理

Visual and Auditory Signal Processing

语音的时频域分析

如何理解语音信号?

如何用更少的属性数据来刻画声音的内容?

对识别类应用---》特征提取

对表示类应用---》表示机理

特征的分析时长

特征的分析时长

➢ **短时分析**

- **语音信号是非平稳信号，但是可以认为10~30ms的时间范围内，语音信号是平稳信号。**
- **短时分析的最基本手段是对语音信号加窗。**

$$x_w(n) = x(m)w(m)$$

常见窗函数: N为窗长

方窗: $w(n) = \begin{cases} 1 & 0 \le n \le N-1 \\ 0 & \text{其它} \end{cases}$

特征的时域分析方法

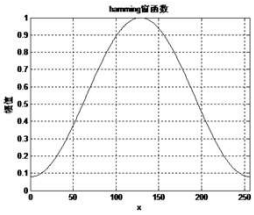
哈明（Hamming）窗

$$w(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)] & 0 \le n \le N-1 \\ 0 & \text{其它} \end{cases}$$

哈宁（Hanning）窗

$$w(n) = \begin{cases} 0.5[1 - \cos(2\pi n/(N-1))] & 0 \le n \le N-1 \\ 0 & \text{其它} \end{cases}$$

特征的时域分析方法



在进行频率分析（FFT）时，让信号具有周期性，消除吉布斯效应的影响

特征的时域分析方法

特征的时域分析方法

- 短时能量、短时平均幅度和短时过零率
- 短时能量

$$E_n = \sum_{m=-\infty}^{\infty}[x(m)w(m)]^2$$

短时能量可用于清浊判决、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。经常是识别系统中特征的一维。

特征的时域分析方法

- 短时平均幅度

$$M_n = \sum_{m=-\infty}^{\infty}|x_w(m)|$$

- 短时过零率：单位时间内过零发生的次数。

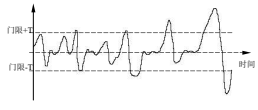
$$Z_n = \sum_{m=-\infty}^{\infty}|\mathrm{sgn}[x(m)]-\mathrm{sgn}[x(m-1)]|w(m)$$

式中

$$\mathrm{sgn}[x(n)]=\begin{cases}1 & x(n)\ge 0\\ -1 & x(n)<0\end{cases} \qquad w(n)=\begin{cases}1/2N & 0\le n\le N-1\\ 0 & \text{其它}\end{cases}$$

特征的时域分析方法

- 短时平均过零率容易受到噪声的干扰，因而提出了门限过零率的思想。



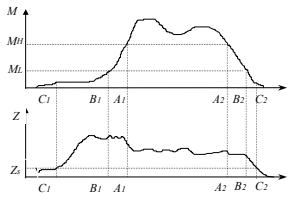
$$Z_n = \sum_{m=-\infty}^{\infty}\{|\mathrm{sgn}[x(m)-T]-\mathrm{sgn}[x(m-1)-T]|+ |\mathrm{sgn}[x(m)+T]-\mathrm{sgn}[x(m-1)+T]|\}w(m)$$

特征的时域分析方法

- 端点检测——能量过零率特征的应用示例
- 对于语音进行“浊音/清音/无声”的判定。
- 在汉语中，若浊音处于音节的末尾，容易通过短时能量来区别，但在音节的前端，清音与环境噪声则很难区分。
- 浊音的能量高于清音，清音的过零率高于无声段。

特征的时域分析方法

- **双门限法**



（1）用较高的短时能量门限M_H确保A1-A2肯定是浊音。

（2）从A1 A2开始向两端搜索，短时能量>较低门限M_l的B1-B2还是语音段

（3）从B1开始向前搜索，短时过零率<门限Zs的为清音部分。

特征的时域分析方法

➢ **短时自相关函数**

- **自相关函数**

对于确定性离散信号 $x(n)$

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

***R(k)*表示一个信号和延迟*k*点后的该信号本身的相似性。**

特征的时域分析方法

自相关函数的性质：

1偶函数： $R(k)=R(-k)$

2 $k=0$**时函数取最大值，对于确定性信号其值为能量。对于随机信号，其值为该信号的平均功率。**

3 如果原序列是周期为*T*的周期信号，那么自相关函数也是周期为*T*的周期函数。

特征的频域分析方法

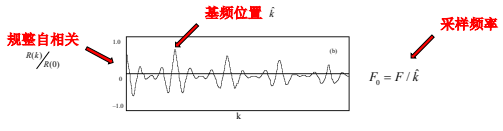
短时自相关函数在假定窗外为0时是偶函数

$$R(k)=\sum_{n=0}^{N-1}x(n)x(n+k)=\sum_{n=0}^{N-k-1}x(n)x(n+k)$$

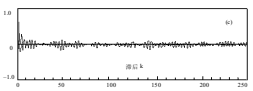
令 $m=n+k$ ，则

$$R(k)=\sum_{m=k}^{N-1}x(m-k)x(m)=\sum_{m=0}^{N-1}x(m)x(m-k)=R(-k)$$

浊音自相关函数波形



清音自相关函数波形



$R(\hat{k})\leq\frac{1}{4}R(0)$

❑ **基音周期检测---短时自相关函数特征的应用示例**

✓基音是指发浊音时声带振动所引起的周期性，它只是准周期性的。

✓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。

✓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 。

✓峰—峰值之间对应的就是基音周期。

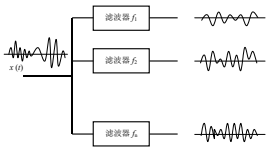
✓在限定K值内的最大峰值出现的位置

特征的频域分析方法

特征的频域分析方法

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

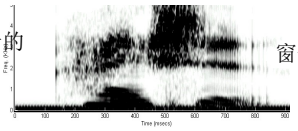
- **滤波器组方法**



特征的频域分析方法

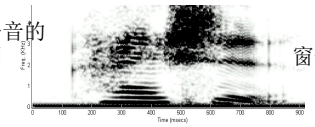
- **语谱图方法**

“开始”语音的宽带语谱图



窗长：6.4ms

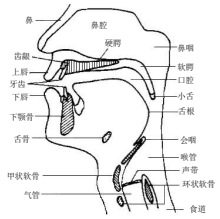
“开始”语音的窄带语谱图



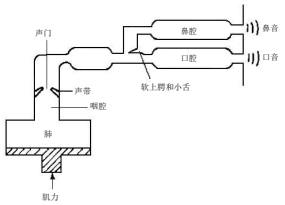
窗长：51.2ms

特征的频域分析方法

- **语音的产生**

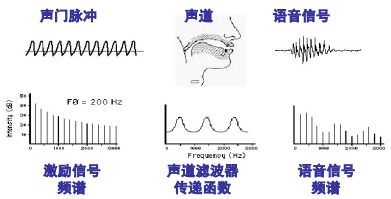


特征的频域分析方法



语音产生的机理图

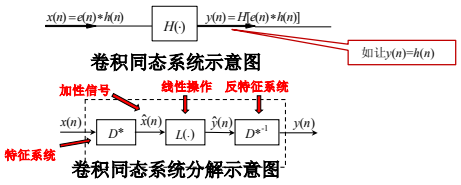
特征的频域分析方法



特征的频域分析方法

- 卷积同态信号处理方法

目的：乘积性组合信号或卷积性组合信号转化为加性信号。将非线性问题转化为线性问题来处理。

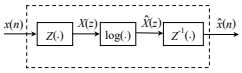


卷积同态系统示意图

卷积同态系统分解示意图

特征的频域分析方法

- 特征系统 D*



第一步是对信号进行Z变换，将卷积信号转变为乘积信号

$$Z[x(n)] = X(z) = E(z) \times H(z)$$

第二步是进行对数运算，将乘积信号变为加性信号

$$\log X(z) = \log E(z) + \log H(z) = \hat{E}(z) + \hat{H}(z) = \hat{X}(z)$$

第三步进行反Z变换运算，变回时域信号

$$Z^{-1}[\hat{X}(z)] = Z^{-1}[\hat{E}(z) + \hat{H}(z)] = \hat{e}(n) + \hat{h}(n) = \hat{x}(n)$$

e(n)已知的条件下很容易通过线性运算得到h(n)

特征的频域分析方法

- 反特征系统 D*-1：它是特征系统的反运算。

- 复倒谱(Complex Cepstrum):将特征系统的输出称为复倒谱或对数复倒谱。

$$\hat{x}(n) = Z^{-1}[\log[Z[x(n)]]]$$

其所在域称之为倒谱域。

特征的频域分析方法

- 倒谱：仅对 $\hat{X}(z)$ 的实部作逆Z变换

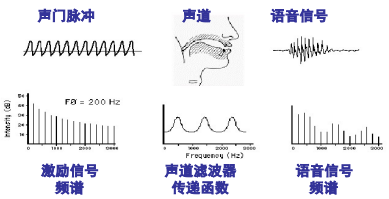
$$c(n) = Z^{-1}[\log|Z[x(n)]|]$$

倒谱不能通过逆特征系统还原成自身。

- 在绝大多数应用场合，特征系统和逆特征系统中的正反Z变换都可以用正反傅立叶变换（DFT和IDFT）来代替。

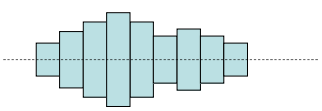
特征的频域分析方法

- 为人类的发声过程建立数学模型



特征的频域分析方法

- 在声学上对均匀的无损耗的管道的声学特性有非常简单的数学描述。均匀：截面积不变；无损耗：不考虑声波在管内的热损耗、粘滞摩擦损耗和管壁的热传导作用。
- 在此基础上，可以将声道简化成一些截面积不等的均匀无损声管的级联。用该模型来逼近真实的声道，称之为声道的时间离散模型。

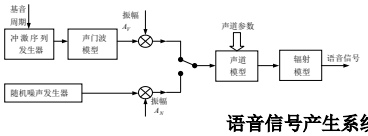


特征的频域分析方法

经过推导，该模型系统的传递函数为如下形式

$$H(z)=\frac{G}{1-\sum_{k=1}^{N}\alpha_k z^{-k}}$$

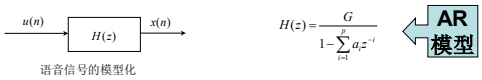
N为级联声管的节数，上式为全极点形式。



语音信号产生系统模型

特征的频域分析方法

- 线性预测（Linear Prediction）分析
- 根据语音信号的产生模型，语音信号$x(n)$可以看作以$u(n)$为激励的一个全极点滤波器的响应。



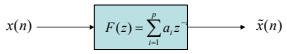
语音信号的模型化

问题：如何在已知$x(n)$的条件下，求出系数$\{a_i\}$ i=1,…, p？

解答：线性预测分析的方法。

特征的频域分析方法

线性预测器：$F(z)=\sum_{i=1}^{p}a_i z^{-i}$



在Z域，是如下乘积关系

$$\tilde{X}(z)=\sum_{i=1}^{p}a_i z^{-i}X(z)$$

反Z变换，可得到如下时域差分方程：

$$\tilde{x}(n)=\sum_{i=1}^{p}a_i x(n-i)$$

从时域角度可以理解为，用信号的前p个样本来预测当前的样本得到预测值

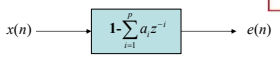
特征的频域分析方法

预测误差

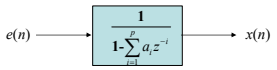
$$e(n)=x(n)-\tilde{x}(n)=x(n)-\sum_{i=1}^{p}a_i x(n-i)$$

可以如下得到该预测器的预测误差

$$E(z)=[1-\sum_{i=1}^{p}a_i z^{-i}]X(z)$$



它的逆滤波器形式：



特征的频域分析方法

- 线性预测器与全极点模型一一对应
- 预测器也不容易确定，系数不同就是不同的预测器，有无数的预测器
- 有一个特殊的预测器：最佳线性预测器
- 最佳线性预测器的预测误差能量最小
- 求最佳线性预测器的过程可以被成为线性预测分析，或者自回归（Autoregressive，AR）分析

特征的频域分析方法

- 思路：在数字信号处理中，一个AR模型与一个最佳的线形预测器是等价的，也就是说，用AR模型的系数a_i构造的预测器必然是最佳预测器，即在最小均方意义上，预测误差能量最小。因此从$x(n)$出发，寻找其最佳预测器，从而得到系数a_i。
- 系数被称为线性预测系数或LPC系数。

特征的频域分析方法

预测误差：

$$e(n)=x(n)-\tilde{x}(n)=x(n)-\sum_{i=1}^{p}a_i x(n-i)$$

短时预测均方误差:

$$E_n=\sum_n e^2(n)=\sum_n[x(n)-\tilde{x}(n)]^2=\sum_n[x(n)-\sum_{i=1}^{p}a_i x(n-i)]^2$$

- 求解过程

使 $\partial E_n/\partial a_k=0,\ (k=1,2,\cdots,p)$ 则有：

$$\frac{\partial E_n}{\partial a_k}=-(2\sum_n x(n)x(n-k)-2\sum_{i=1}^{p}a_i\sum_n x(n-k)x(n-i))$$

特征的频域分析方法

得到线性方程组

$$\sum_n x(n)x(n-k)=\sum_{i=1}^{p}a_i\sum_n x(n-k)x(n-i)\qquad k=1,2,\cdots,p$$

若定义 $\Phi(k,i)=\sum_n x(n-k)x(n-i)\qquad k=1,2,\cdots,p\qquad i=0,1,2,\cdots,p$

则方程组可简写为 $\sum_{i=1}^{p}a_i\Phi(k,i)=\Phi(k,0)$

一个由p个方程组成的有p个未知数的线性方程组

求解方程，可得到LPC系数

特征的频域分析方法

要构造信号的AR模型，还应估算增益因子

AR模型的差分方程形式 $x(n)=\sum_{i=1}^{p}a_i x(n-k)+Gu(n)$

因此可计算预测误差 $e(n)=\sum_n x(n)-\tilde{x}(n)=Gu(n)$

且 $E_n=G^2\sum_n u^2(n)$

激励信号$u(n)$总能量可以认为近似为1,因此有 $\hat{G}=\hat{E}_n^{1/2}$

特征的频域分析方法

- 求LPC系数需考虑两个因素

（1）模型阶数的选择 $p=2D+1$，D 是共振峰的个数

（2）考虑口唇的高频衰减特性，在线性预测分析之前，需要通过预加重进行高频提升 $1-\alpha z^{-1}$

特征的频域分析方法

- 自相关法

我们定义 $\Phi(k,i)=\sum x(n-k)x(n-i)$ 时，未将求和范围具体化。一种较直接的方法是，认为语音段外的数据全为零，只计算范围n以内（$0\leqslant n\leqslant N$）的语音数据。

$$\Phi(k,i)=\sum_{n=0}^{N-1+p}x_w(n-k)x_w(n-i)\qquad \begin{matrix}k=1,2,\cdots,p\\ i=0,1,2,\cdots,p\end{matrix}$$

或

$$\Phi(k,i)=\sum_{n=0}^{N-1-(k-i)}x_w(n)x_w(n+k-i)\qquad \begin{matrix}k=1,2,\cdots,p\\ i=0,1,2,\cdots,p\end{matrix}$$

$x_w(n)$为加窗后的语音数据。

特征的频域分析方法

由于短时自相关函数可以表示为：

$$R_n(k)=\sum_{n=0}^{N-1-k}x_w(n)x_w(n+k)$$

且有 $R_n(-k)=R_n(k)$

则 $\Phi(k,i)$ 可以表示为

$$\Phi(k,i)=R_n(k-i)=R_n(|k-i|)\qquad \begin{matrix}k=1,2,\cdots,p\\ i=0,1,2,\cdots,p\end{matrix}$$

求解LPC系数的方程组就可以写为：

$$\sum_{i=1}^{p}R_n(|k-i|)\hat{a}_i=R_n(k)\qquad k=1,2,\cdots,p$$

特征的频域分析方法

将其转换成矩阵形式

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \dots \\ \dots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ \dots \\ R_n(p) \end{bmatrix}$$

这种方程为Yule-Walker方程，其系数矩阵被称为托布里兹(Toeplitz)矩阵。具有如下性质：

（1）$p \times p$阶的对称阵。

（2）沿着主对角线及任何一条与主对角线平行的斜线上的所有元素都相等 。

特征的频域分析方法

Yule-Walker方程可以用递推的方式来求解。典型的方法有：

- 莱文逊—杜宾（Levinson—Durbin）递推算法
- 舒尔（Schur）递推算法

特征的频域分析方法

莱文逊—杜宾(levinson-durbin)递推算法

- 不直接计算p阶预测器
- 从一阶预测器开始，逐一递推各阶预测器
- 第i阶预测器的系数可以用第$i-1$阶预测器的系数递推得到
- 直到递推出p阶预测器的系数
- 用到了i阶预测器的预测误差能量 $E^{(i)}$和一个中间系数k_i

张贤达等,《现代信号处理》,清华大学出版社

特征的频域分析方法

(1)计算自相关系数 $R_n(j), \quad j=0,1,\dots,p$

(2)初值 $E^{(0)}=R_n(0) \quad i=1$

(3)开始按如下公式进行递推运算

$$k_i=\frac{R_n(i)-\sum_{j=1}^{i-1}a_j^{(i-1)}R_n(i-j)}{E^{(i-1)}}$$

$$a_i^{(i)}=k_i$$

$$a_j^{(i)}=a_j^{(i-1)}-k_i a_{i-j}^{(i-1)} \qquad j=1,\dots,i-1$$

$$E^{(i)}=(1-k_i^2)E^{(i-1)}$$

(4)$i=i+1$。若$i>p$则算法结束退出，否则返回第（3）步，

特征的频域分析方法

经过递推计算后,最终解为:

$$\hat{a}_j=a_j^{(p)}, \qquad j=1,2,\dots,p \qquad E^{(p)}=R_n(0)\prod_{i=1}^{p}(1-k_i^2)$$

可以推知

$$|k_i|\le 1, \qquad i=1,2,\dots,p$$

k_i称为反射系数，也称PARCOR系数。

特征的频域分析方法

- 协方差法：

重新定义求和范围

$$\Phi(k,i)=\sum_{n=0}^{N-1}x(n-k)x(n-i) \qquad k=1,2,\cdots,p \qquad i=0,1,2,\cdots,p$$

设 $(n-i)=m$

$$\Phi(k,i)=\sum_{m=-i}^{N-i-1}x(m+(i-k))x(m) \qquad k=1,2,\cdots,p \qquad i=0,1,2,\cdots,p$$

此时不再满足 $\Phi(i+1,k+1)=\Phi(i,k)$,因而系数矩阵变成如下形式

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & \Phi(1,3) & \dots & \Phi(1,p) \\ \Phi(2,1) & \Phi(2,2) & \Phi(2,3) & \dots & \Phi(2,p) \\ \Phi(3,1) & \Phi(3,2) & \Phi(3,3) & \dots & \Phi(3,p) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \Phi(p,1) & \Phi(p,2) & \Phi(p,3) & \dots & \Phi(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \dots \\ \dots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \Phi(1,0) \\ \Phi(2,0) \\ \Phi(3,0) \\ \dots \\ \dots \\ \Phi(p,0) \end{bmatrix}$$

特征的频域分析方法

此时的系数矩阵不再是一个托布里兹矩阵，它一般用乔里斯基（Choleskey）分解法来求解。

- 自相关法和协方差法的比较

➢自相关法必须对语音信号进行加窗处理，规定了信号的长度范围，假定窗外的语音样本值为零，所以自相关法误差较大，计算结果精度差，但自相关法能够保证系统的稳定性 。

➢协方差法因不需要加窗，所给出的参数估值要比自相关法精确的多，但不如自相关法稳定，另外乔里斯基分解法因没有快速算法，也需要较大的计算量 。

特征的频域分析方法

LPC倒谱系数（LPCC）

- 倒谱是通过对信号进行Z变换，取对数，再反Z变换来得到的。
- 求单位冲激响应$h(n)$的倒谱 $\hat{h}(n)=Z^{-1}[\log H(z)]$
- 它也反映了信号的谱包络信息。

有：
$$H(z)=\frac{1}{1-\sum_{i=1}^{p}a_i z^{-i}}$$

$\hat{H}(z)=\log H(z)$ 可以展开成级数形式 $\hat{H}(z)=\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}$

$$\log\left[\frac{1}{1-\sum_{i=1}^{p}a_i z^{-i}}\right]=\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}$$

特征的频域分析方法

将上式两边同时对 z^{-1} 求导

$$\frac{\partial}{\partial z^{-1}}\log\left[\frac{1}{1-\sum_{i=1}^{p}a_i z^{-i}}\right]=\frac{\partial}{\partial z^{-1}}\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}$$

有

$$\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\frac{\sum_{i=1}^{p}ia_i z^{-i+1}}{1-\sum_{i=1}^{p}a_i z^{-i}}$$

$$(1-\sum_{i=1}^{p}a_i z^{-i})\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\sum_{i=1}^{p}ia_i z^{-i+1}$$

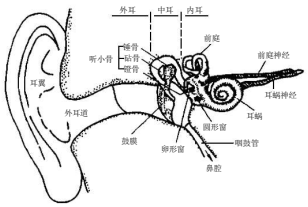
特征的频域分析方法

得到 $\hat{h}(n)$和 a_i 间的递推关系为

$$\begin{cases}\hat{h}(1)=a_1\\ \hat{h}(n)=a_n+\sum_{i=1}^{n-1}(1-\frac{i}{n})a_i\hat{h}(n-i), & 1\le n\le p\\ \hat{h}(n)=\sum_{i=1}^{p}(1-\frac{i}{n})a_i\hat{h}(n-i), & n>p\end{cases}$$

特征的频域分析方法

感知机理的仿真



特征的频域分析方法

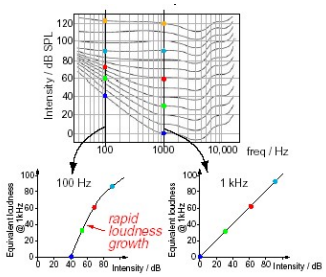
- 正常人耳能感知的频率范围为16.4Hz~16KHz;强度范围为0dB~120dB。
- 音调是人耳对不同频率声音的一种主观感觉。单位为Mel，与频率近似的满足方程：

$$p_{Mel}\cong(1000/\lg 2)\times\lg(1+0.001f_{Hz})$$

- 响度用来描述人耳对不同频率的纯音的辨别灵敏度。单位为Phon。1Phon等于1kHz纯音的1db声强级。为了确定一个音的响度，需要调节1kHz纯音的声强，使其与目标音一样响，此时的声强就是待求响度。

特征的频域分析方法

- 等响度曲线



- 掩蔽效应

特征的频域分析方法

- **Mel频率倒谱系数（MFCC）**

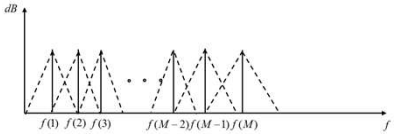
  人的耳蜗实质上的作用相当于一个滤波器组，耳蜗的滤波作用是在对数频率尺度上进行的，在1000Hz以下为线性尺度，而1000Hz以上为对数尺度，这就使得人耳对低频信号比对高频信号更敏感。

  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。

（1）将时域信号$x(n)$后补若干以形成长为N（一般取N=512）的序列，然后经过FFT变换的线性频谱$X(k)$.

特征的频域分析方法

（2）将线性频谱$X(k)$通过Mel频率滤波器组得到Mel频谱。



(3) 对每个滤波器的输出信号取对数能量。

（4）对这组对数能量值做DCT变换。

特征的频域分析方法