谱频谱及一些与其相类似的函数

A. M. Noll

（美国华盛顿总统办事机构科技办公室）

一、周 期 信 号

1．模型

假设本文所讨论的周期信号可以表示为时变线性滤波器的输出，滤波器以具有时变周期的准周期信号作为输入。假设滤波器的脉冲响应h(t) 在几个周期内随着时间的变化很小，而激励信号或源信号的周期e(t)在几个周期上是恒定的，线性滤波器的共振模即是传递函数H(ω)的峰值。测定基频的问题是：给定信号f(t)，测定激励信号的周期。

2．自相关分析的理论

如前所述，假定一个周期信号在整个分析区间内周期恒定。如果无限地重复这个周期信号，就会形成一个周期信号，这个信号的频谱由频率轴上相隔一个基频的许多δ函数组成。在$|t|\leqslant T$时，具有$w(t)$加权，而在别处为0的一个有限分析区间的效应是$w(t)$的付里叶变换和脉冲串的卷积。其结果是一个具有谱波纹的“周期”谱。

为了防止把时间函数的通常的频率分量和对数谱中的波纹“频率”混淆起来，Bogert和Tukey使用了解释词“谱频率”（quefrency）来描述谱波纹的“频率”。谱频率的单位是周/赫或简单地用秒。同样地，这个特殊波纹的“周期”称作谱周期（repiod）。

因为确定一个周期信号的周期的常用方法是求它的付里叶变换，功率谱的付里叶变换应有一个与功率谱的谱周期相对应的峰。这个峰将发生在一个等于原信号周期的谱频率处。所有这些也许听起来奇怪，但是这只不过是自相关函数的直观的推广，而自相关函数是根据在功频谱中测出的谱周期性来建立的。

3．譜频譜分析的理论

激励信号$s(t)$是周期性的信号源，这个信号通过滤波器后所得到的最后信号$f(t)$是激励信号$s(t)$*和滤波器脉冲响应$h(t)$的卷积：$f(t)=s(t)*h(t)$。但是另一方面，$f(t)$的功率谱$|F(\omega)|^2$等于$s(t)$和$h(t)$的功率谱的乘积：$|U(\omega)|^2=|S(\omega)|^2\cdot|H(\omega)|^2$。于是，自相关函数即是$|S(\omega)|^2$和$|H(\omega)|^2$的付里叶反变换的卷积。就许多信号而言，传递函数$H(\omega)$是这样的，

---

*原文误为e(t)。—译注

它使求得的卷积给出多重峰值。实际上，求卷积反而乱了套！然而，如果取功率谱的对数，这二个谱就变成可以相加的：$\log|U(\omega)|=\log|S(\omega)|+\log|H(\omega)|$。图1所示的是滤波器脉冲响应的对数谱在信号谱中引起低的谱频率的波纹。激励信号的对数谱使信号谱具有谱周期性的精细结构。所以，对数谱的功率谱是由一个峰值和低的谱频率成份组成；这个峰值位于与谱周期性精细结构相对应的谱频率上，而低的谱频率成份是与滤波器的响应相对应的。这二个效应是相加的，同时，因为它们占据不同的谱频率范围，所以，它们是可以区分的。这样，只要取功率谱的对数，就导出了一个新函数。再次借用Bogert和Tukey的术语，这个新函数称作谱频谱（cepstrum），并正式定义为对数功率谱的功率谱：

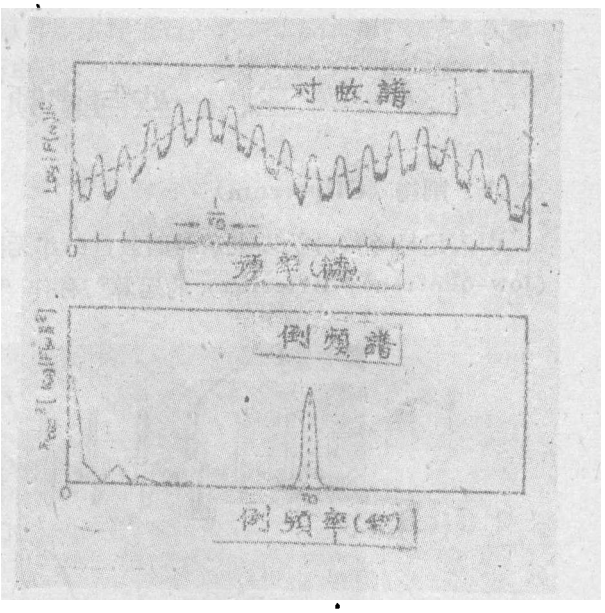


图 1*

$$C(\tau)=|F[\log|U(\omega)|^2]|^2=|F_{\cos}[\log|U(\omega)|^2]|^2$$

式中，F和$F_{\cos}$分别表示复付里叶变换和付里叶余弦变换。

二、谱 频 谱

谱频谱$C(\tau)$已正式定义为对数功率谱的功率谱。因为对数功率谱是一个偶函数，所以这个定义等效于对数功率谱的余弦变换的平方：

$$C(\tau)\equiv\left\{\int_0^\infty \log|U(\omega)|^2\cos(\omega\tau)d\omega\right\}^2$$

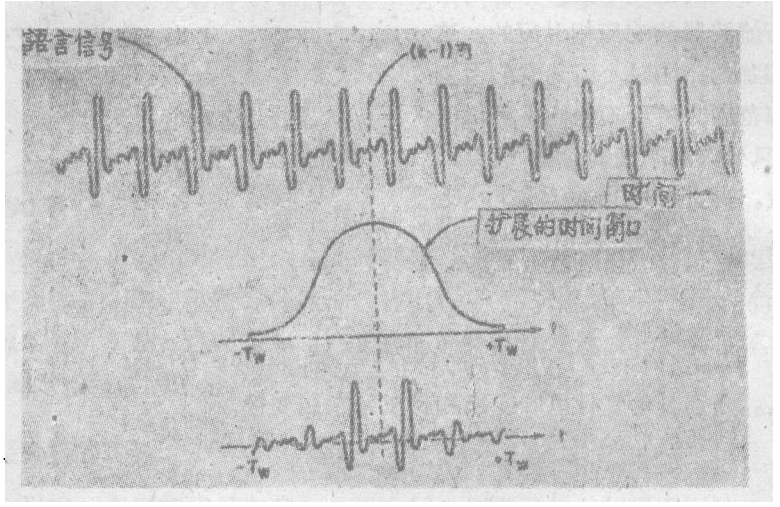
许多信号的基频随时间而变。所以，需要各个短时间内信号段的一系列谱频谱。这可用一个在某有限时间区间以外是零的一个函数乘该时间信号来完成。这个函数犹如一个窗口，通过这个窗口就可以看到信号，它的作用将在后面作更详细的讨论。如图2所示，对时限信号作谱分析，得到对数谱，然后再求出谱频谱。随后新的一段时间信号进入窗口，作了同样的分析便求得另一个谱频谱。当重复进行这个过程时，就得到一系列瞬时谱频谱。如需要的话，此时间窗口也可以观察信号的重叠部分。

*本文图中倒频谱和倒频率分别改为谱频谱和谱频率。一译注

三、一些与谱频谱相类似的函数

1．削譜（Clipstrum）

引入记号〈f(t)〉以表示对函数f(t)作无限削峰的运算。用下标LQ表示滤掉低谱频(low-quefrency liftering)的运算*。称作“削谱”的新函数因把所有的峰削去而得名。



第k段瞬时频谱：

$$U_k(\omega) \equiv \int_{-T_w}^{T_w} [X(t)] e^{-j\omega t}\, dt \quad (X(t)\text{是图中第 3 个波形})$$

第k段对数功率谱：

$$\log |U_k(\omega)|^2 = \log[Re^2 U_k(\omega) + Im^2 U_k(\omega)]$$

第k段瞬时谱频谱：

$$C_k(q) \equiv \left| \int_0^{\omega_c} \log \left| U_k(\omega) \right|^2 \cos \omega q \, dq \right|^2$$

图　2

削谱可按照下列运算而得到：

（1）把时间信号$f(t)$无限削峰：

$$g(t) \equiv \langle f(t) \rangle;$$

（2）计算$g(t)$的对数功率谱：

$$L(\omega) \equiv \log |F[\langle f(t) \rangle]|^2;$$

---

*作者在另一文(J. A. S. A 44, 1585—1991; 1968)中，亦称高通滤谱器（high-pass lifter）。—译注

（3）用滤谱器（*lifter*）在频率域中对$L(\omega)$进行低谱频滤谱。该滤谱器在整个600赫频率区间内具有*hamming*脉冲响应。这个运算可以这样进行：首先求*hamming*频率函数和对数功率谱的卷积，得到一个平滑化的对数功率谱，然后，从对数功率谱中减去这个平滑化的对数功率谱。代数表示式为：

$$L_{LQ}(\omega)=L(\omega)-L(\omega)*h(\omega)$$

式中

$$h(\omega)=\begin{cases}0.54+0.46\cos(\omega/600)\ ; & |\omega|\leqslant 600\pi\\ 0\ ; & |\omega|>600\pi\end{cases}$$

（4）把已滤谱的对数谱作无限削峰：

$$I(\omega)\equiv\langle L_{LQ}(\omega)\rangle$$

（5）计算$I(\omega)$的付里叶余弦变换并求平方，于是求得最后结果：

$$C(\tau)\equiv\{F_{\cos}[I(\omega)]\}^2$$

把上述运算放入一个方程式中，就得到$f(t)$的削谱的表示式：

$$C(\tau)\equiv\{F_{\cos}[\langle\log|F[\langle f(t)\rangle]|^2-\log|F[\langle f(t)\rangle]|^2*h(\omega)\rangle]\}^2$$

式中，F表示复付里叶变换，而$F_{\cos}$是余弦变换。

乍一想，所有这些似乎非常奇怪，实际上，我们确有充分的根据指望削谱象基频检波器那样相当好地工作。无限削峰不能消除频谱中的谐波的精细结构，相反地，通常引入附加的谐波。所以一个无限削峰信号的对数功率谱仍应具有一频率，这一频率位于等价于基频的二谐波之间。然而，对数谱可以仅由正项组成，并且还具有一些低谱频分量，因此需要进行滤谱，以便仅剩下谐波的精细结构。于是，已滤谱的谱也可以无限削峰，它的付里叶余弦变换应该具有一个与位于已削峰的、被滤谱的谱中的谐波相一致的峰。

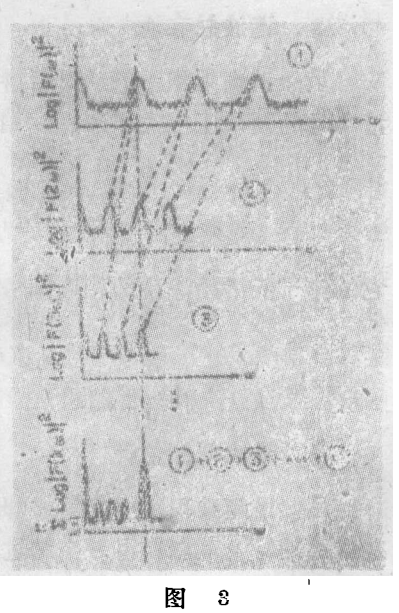


图 3

2．谐频乘积谱

在对数频谱中的许多谐波峰就是几倍基频的谐波，所以，若把对数谱作整数倍的压缩，且将结果相加，那么，将在所有谐波的重合处得到一个大峰。如图3所示。在数学上讲：

$$\log\pi(\omega)\equiv\sum_{k=1}^{K}\log|U(k\omega)|^2=\log\prod_{k=1}^{K}|U(k\omega)|^2$$

取反对数后就导得谐波乘积谱

$$\pi(\omega)\equiv\prod_{k=1}^{K}|U(k\omega)|^2$$

这种方法的直观解释为：对数谱中的谱波峰相干叠加，而其他部份是不相关的，并非相干叠加。频率压缩产生了最后的大尖峰。

当频率压缩形成谐频乘积谱时，因为对数谱的低谱频结构只是增加了些模糊，因此要把这些低谱频消除掉。为此，在计算乘积谱之前，用一个*hamming*函数平滑该谱，然后从原来的谱中减去平滑了的谱。如果以W（ω）表示与*hamming*时间窗口相应的频率域，即：

$$W(\omega)=\begin{cases}0.54+0.46\cos\omega/2B; & |\omega|\leqslant 2\pi B\\ 0; & |\omega|\geqslant 2\pi B\end{cases}$$

那么已作了低谱频滤谱的对数谱$L_{LQ}(\omega)$为：

$$L_{LQ}(\omega)=\log|F(\omega)|^2-W(\omega)*\log|U(\omega)|^2$$

因此諧频乘积谱为：

$$\pi_{LQ}(\omega)=\prod_{k=1}^{K}L_{LQ}(k\omega)$$

3．谐波总和譜

虽然谐频乘积谱是从对数谱直观地导出的，但利用对数谱并不是唯一的途径。对于功率谱本身的频率压缩式样来说，相干求和论述也应成立。由此，谐波总和谱可定义为：

$$\sigma(\omega)\equiv\sum_{k=1}^{K}|U(k\omega)|^2$$

它的作了低谱频滤谱的相应部份为：

$$\sigma_{LQ}(\omega)\equiv\sum_{k=1}^{K}\log^{-1}[L_{LQ}(k\omega)]$$

4．最大似然估计

虽然谐频乘积谱中的许多峰会因尖锐化而激化起来，但最为重要的事实是：在基频（或基本周期）上可以获得一个超过某一固定门限的峰。对此，所有三种办法，即：谐频谱，谐频乘积谱和谐波总和谱都可令人满意地达到要求。然而，对基频测定显然应有一种“最佳”方法，而标准信号处理技术应当给出这种方法。这促使造成对周期信号周期的最大似然估计的研究。David Slepian已对周期信号的最大似然估计作了数学推导，下面就是该技术的直观证明。

考察一段长度为T，周期为τ_0的周期信号r(t)（见图4）。尽管τ_0未知，但可以把信号分成若干段，每段长度为τ，并把它们加起来，就得到一个新函数；

$$R(t,\tau)\equiv\begin{cases}\dfrac{1}{N+1}\sum_{n=0}^{N}r(t+n\tau), & 0\leqslant t<b\\ \dfrac{1}{N}\sum_{n=0}^{N-1}r(t+n\tau), & b\leqslant t<\tau\end{cases}$$

其中$T=N\tau+b$。然后，构成$R(t,\tau)$平方的积分，这一积分在不同的积分区间用该区间求和的总数作加权：

$$J(\tau)\equiv(N+1)\int_0^b R^2(t,\tau)\,dt+N\int_b^\tau R^2(t,\tau)\,dt$$

当选择τ等于音调周期τ_0时，所有小区间就相干叠加，所以$J(\tau)$在$\tau=\tau_0$时最大。使$J(\tau)$为最大的τ值就是周期的最大似然估计值。

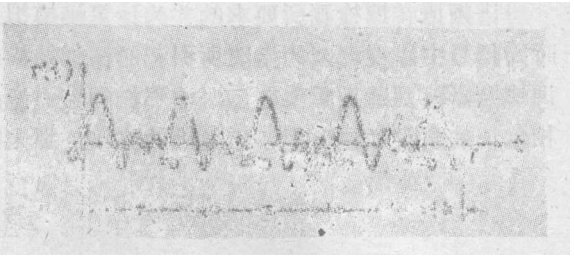


图4　一段长度为T，周期τ_0未知的周期性语言信号；它分成长度为τ的N个小段，使$T=N\tau+b$

四、音调的测定

空气经过声带而产生的周期性喷气声在声道中发生谐振时，就产生有声的语言。为了测定音调周期，需从被观察的语言信号获得信号源的周期性。而且，判定是有声语言或无声语言需要精确测定源信号中这种周期性喷气声存在与否。这个似乎简单的问题在过去几十年中成了大量研究的课题。一种精确的音调检测器，除了从纯粹的研究观点来看，在语言声的分析中有明显的用途外，它也必定能作为大多数语言带宽压缩方案中的一个整体部件。因此一种能很好地工作于有限带宽的杂乱语言信号情况下的精确音调检测器的设计仍然是语言处理研究中疑难问题之一。

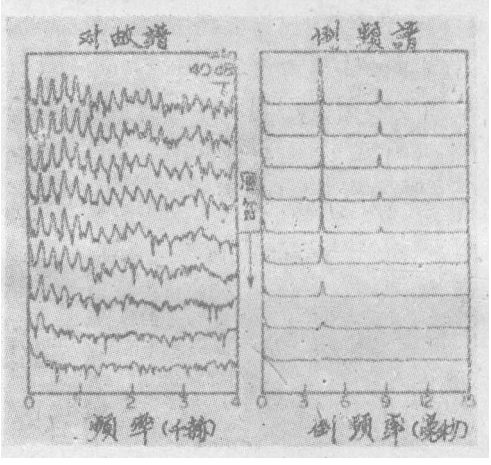


图5　用优质电容微音器记录的一段女声语言，左图是用计算机计算的瞬时对数谱，右图是相应的谱频谱。在瞬时对数谱中，横坐标为频率，纵座标为对数幅值，每条连续曲线为40毫秒的一段语言的对数谱，这些语言段之间有一部份是重叠的，每隔10毫秒向下划一条。用40毫秒长的 hamming 时间窗口进行观察。

测定周期信号的基频应是一个比较简单的问题。但遗憾的是由于一系列的原因，对语言信号来说并不如此。从男声80赫到女声400赫的范围内，音调频率都能覆盖。当某一个人作一次发言时，这些频率会变化很大，实际上，音调周期的突然加倍是不罕见的。假如音调检测器在电话音色的语言条件下运用，那末它一定是在没有基频的信号条件下工作的。同时可以预期会有相位漂移、弱的信噪比和非线性失真。因此，十分简单的音调测定问题就成为十分困难的信号音调测定问题，因为这种信号的先验统计特性是未知的，在通过非线性通讯网络后，它将淹没在噪声中。

一个实际的音调检测器必须能在缺少基频或者谐波随频率衰减并且只有几个谐波是有效的情况下工作。所有使用语言时不可能估计出这类变化的，所以就引入了人工计算机发生的信号，这种信号是由湮没在计算机计算的白噪声中的许多正弦波之和组成的。用同一段白

噪声，这样因随机选用噪声段而引起的任何起伏就可消除。这些正弦波的频率和幅度可予以变化的，以致可以在谱中形成消除谐波和衰减振幅的一系列组合。

谱频谱音调检测所要求的并为计算机模拟所证实的一些优点，首先是，基频分量不须在时间信号中出现，因为谐波所引起的精细结构或谱的波纹产生谱频谱的峰值。因此谱频谱音调检测器特别适用于象电话语言那样的带通滤波信号。因为只利用了功率谱，所以完全忽略相位。如果附加白噪声不破坏谱的波纹，那末附加的白噪声是不会衰减太快的。

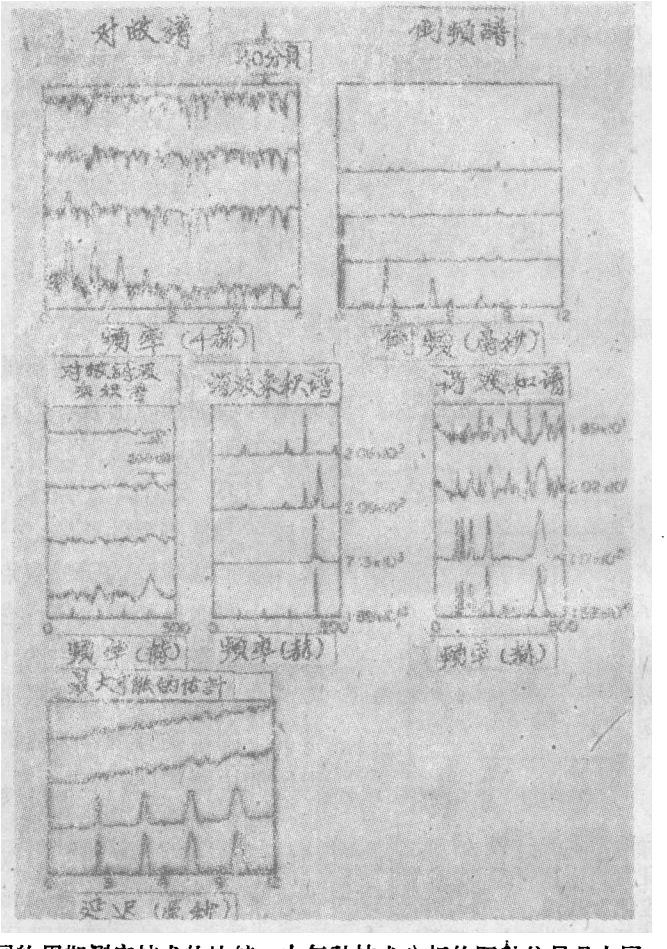


图 6　不同的周期测定技术的比较。由每种技术分析的四种信号是由同一段白噪声的四种不同级和五个正弦信号组成，这些信号的幅度以 6 分贝/倍频程衰减。这四种噪声级按12分贝步级减小，而最上面一条曲线的噪声平均谱级差不多等于基频峰值。

看来有一点是绝不会令人满意的，即谱频谱音调检测对于扰乱很大的语言信号是不适用的，所以，需要重新开始寻找“理想的”音调检测器。这种想法就引出了削谱作为简化硬设备的一种尝试；又引出了谐波总和谱和谐频乘积谱，以作为改善处理嘈杂的语言信号特性的一种尝试。

一般的结论是：除了最大似然估计会得到幅度增加的多重峰外，谐频乘积谱、谐波总和谱以及最大似然估计这三种方法的结果大约是同样好。将谱频谱法与其他方法相比较，对谱频谱法来说直到噪声级再下降12分贝后，才能清楚地辨认谱频谱的峰。换言之，对检测和测

定由淹沒在附加白噪声中的具有谐波关系的正弦波组成的信号的基频来讲，其他方法要比谱频谱法好。

尽管对各种方法评价时，谱频谱法用于淹沒于白噪声中的正弦波信号时有某种缺点，但是在非嘈杂语言的音调测定时，谱频谱法仍然“优于”其他方法。实际上，谱频谱法已测得了孤立地拍击声带的音调；也已正确指明，在鼻音语言段的末端，音调周期是加倍的。而对同一信号，用其他方法测定其音调周期，要么测不出，要么测不准。诚然，对于处理非常杂乱的信号，其他方法，特别是谐频乘积谱法，要更好些，并且除音调测定外还能在别的場合得到应用。

（毛汝夏 胡嘉忠译 丁达夫 冯绍松校）

参 考 文 献

1. Bogert, B.P., Healy, M.J.R. and Tukey, J.W(1963). "The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking", (M.Rosenblatt, Ed.) "Proceedings of the Symposium on Time Series Analysis", 209-243.
2. Cooley, J.W. and Tukey, J.W. (1964). "An algorithm for the machine calculation of complex Fourier series", Math. Comp. 19, 297-301.
3. Kincaid, Thomas G. and Scudder, Henry J.111, (1967). "Estimation of periodic signals in noise." J.A. S. A. 42, №, 5, 1166.
4. Miller, R.L. (1968). "Pitch determination by measurement of harmonics." J. A. S. A. 44, №1, 390.
5. Noll, A. Michael(1964). "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection." J. A. S. A. 36, №2, 296-302.
6. Noll, A. Michael(1967). "Cepstrum pitch determination" J.A.S.A. 41, №2, 293-309.
7. Noll, A. Michael(1968). "Clipstrum pitch determination." J.A.S.A.44, №6, 1585-1591.
8. Noll, A. Michael. "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate." "Proceedings of the Symposium on Computer Processing in Communications", Vol. XIX, 779—797.