专题技术讲座

数字信息传輸

中国人民解放军国防科学技术大学情报资料研究室

一九八〇年六月

目　　录

1. 语音编码 ……1
2. 用纠错编码作为数据压缩的码字法和伴随式法 ……35
3. 黑白电视信号低比特率的差分 PCM ……43
4. 阿波罗数字指令系统 ……56
5. 有限几何 ……67

语　音　编　码

1. 语音数字编码

本文范围

当数字技术的发展及大规模集成在经济上取得成就的时候，人们的兴趣重新集中到有效的语音数字编码及传输方法上。根本目的仍然是使语音传输质量尽可能高、所需信道容量尽可能低且成本最低。但由于数字硬件经济性的指望，用新的先进数字方法来达到这个目的的想法却是新颖的。

典型地说，语音编码的成本必定与编码器的复杂程度有关，而且其复杂程度同样必定与编码效率及信道利用率有关。所以，复杂的（及可能有效的）数字编码常为高成本的恶果所阻碍。但是，器件集成化的进一步发展引人注目地使这个趋势发生变化。因而要及时地深入研究语音信号的特性，使其特别适宜于用数字表述。因此，本文的目的是概述关于“数字语音”的现有的知识及效能，并设想近期内新发展的前景。我们的观点难免是狭隘的，因为谈及由我们所具有的第一手感性材料而得到的结果是比较容易的，希望读者原谅这种偏见。

任何传输系统的整体设计都需要（在许多方面）对信号质量、传输比特率及编码器成本等因素的配合进行最佳选择。这种适当的选择很大程度上取决于传输环境（例如地面电线、玻璃纤維或无线电）。总系统的最佳化问题大大地超出了本文范围。但是，我们将尽力指出编码设计所不可避免的系统问题的联系，如传输误码率、多级编码（串联）、变速编码、分组传输及加密。

保真标准

任何信号质量的评定实际上是用保真度来计量。对大多数通信系统而言，定量的测量是有困难的，因为这要涉及到人们的知觉作用。语音质量习惯上用听懂在“说什么”和“谁说的”这种标准来评定。人们积极探求准确反映这些因素的客观计量方法，但却难于用一般原则确定。

如不考虑论述的片面，则有些方法可用来对语音质量作定量测试。它们大都是由单字清晰度测试、谈话者识别测试及信号掩模（masking）试验所得出来的。所有这些都由人来收听。在语音编码器的设计中，常用这些数据作指标。此外，短时间频谱的计量和加权信噪比、精细的判认、对无从捉摸语音的描述等级、感觉的客观定量估计等，这些问题将随着讨论的展开而列入专题评论。

波形编码器与信源编码器（声码器）

语音编码器的一大类称为波形编码器。正如其名称所表示的那样，波形编码器实质

上是力求逼真再现信号波形。原则上，这些编码器被设计成与信号无关。因此它能对各式各样的信号，如语音、音乐、单音及音频频带数据等同等地进行编码。也能应付谈话者的种种特性及噪声环境。为要維持这些最少复杂的优点，波形编码器旨在中等地节省传输比特率。

波形编码器可以达到最佳化且对多数专用信号可以得到较高的编码效率。典型地是通过观察给定信号的统计特性而得到的，因此波形编码器对这类信号（即语音）具有最小的编码误差。因此它不是按照信号某些物理模型来使语音信息参数化，而是根据语音波形的统计特性进行加工的。

第二类语音编码器是利用关于信号在信源中如何形成的先验知识简要描述为依据。基本概念是信号的某些物理条件可以被度量，而且有利于有效地描述信号。这就意味着信号必须与具体的模型（在这里指专用语音）相适应，并相应地参量化。我们称这种利用信号形成约束的编码技术为“信源编码”。语音信源编码通常叫做声码器(Vocoder，即Voice coder 的缩写)。

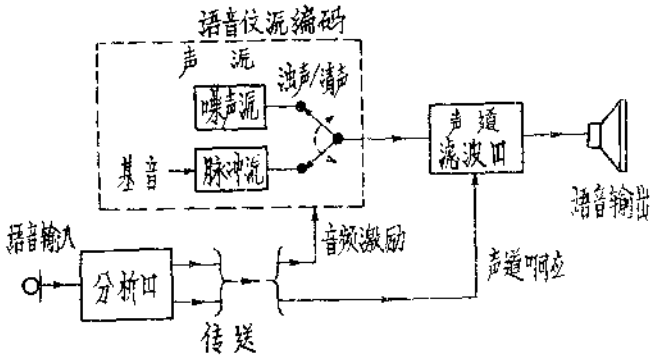


图 1 语音形成的信源系统模型

传统的语音形成模型，即从所谓“通道声码器”开始的模型是如图 1 所示的信源系统模型。声音形成器（信源）被假设为与该系统的消息调制。音域滤波是线性可分的。另外假设语声是浊音或清音中任一种，它们是由准周期的声带声或由湍流空气流动而产生的随机声音。该模型的参数（音频基音、调制滤波器的极点频率及相应的振幅参数）将在后面讨论。这里，要尽量注意到，通过对参数的高度细心的调整，可证实所复现的信号将是高质量的。更普遍地，在实际的一次分析/合成的传输中，声码器往往是不可靠的(如区别浊音/清音及基音值等参数)、性能与谈话者密切相关且输出语音只有人造的（比原来的较低）质量。这些特点是声码器可以达到的最好性能。不过由于利用其信号参数，声码器却能大大地节省传输带宽。

在波形编码器与声码器之间可以认为有一种中间范畴，其设计准则既非保持波形也非信号模型化，而是以听觉佳美的情况下保持语音信号的短时间振幅谱为指导。这个中间范畴可以兼顾波形和信源编码器二者的一些优点。

语音编码的传输速率

图 2 示出了一个普遍关心的语音编码传输速率谱。图中突出表示出非专用语音波形编码器相对地需要较高的传输速率，而专用语音的声码器数字化用相对较低的比特

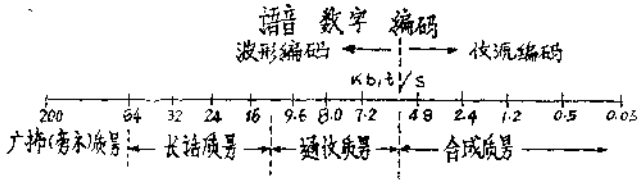


图 2 语音编码传输比特率（非线性刻度）及相应质量谱

率。图中也表明了在规定的比特率上目前能达到的语音复现的质量。质量特性用旁示、长(途)(电)话、通信及合成等质量表示。

我们用长话这个术语有点不严谨，只是为了与模拟语音信号比较，模拟语音信号一般具有如下性能：[频率范围为 200～3200 HZ；信噪比大于等于 30 dB；谐波失真小于等于2.3%]。如图 2 所提示，我们可以很快了解如何构成数字编码器，使其以 16kbit/s或高一点的编码速率达到语音电话的长话质量。到目前为止，我们已经知道即使用任意复杂的方法也无法用比 16kbit/s 低很多的速率去达到长话质量。

当速率超过 64kbit/s 时，对于有效带宽较正常电话宽得多(即 0～7 KHZ 或更宽)的输入信号可能达到长话质量要求的信噪比及谐波失真。这种质量等级一般可称为旁示质量，它适用于许多不同无线电广播设备的数字化。

在速率低于 16 kbit/s 和特定的 9.6～7.2 kbit/s 数据速率范围内，我们知道怎样用波形编码器达到语音通信质量。此时信号大多是可懂的，但质量显著下降，有某些可检测的失真及谈话者识别能力多半要降低。另外，电路复杂程度是传输速率的函数。

在信源编码范围内的编码器（声码器），速率为 4.8 kbit/s 或较低时可达到合成质量，这时信号一般已失去基本的天然性，典型地声音细长且有时自动相似。谈话者识别能力显著地降低，编码器特性随谈话者而变。

在本文之余，我们给出波形编码器与声码器的特性概述。我们将进一步不很严格地讨论两种编码器的具体的常用例子。可能情况下我们将通过适当的引证尽可能表明初期的工作和有关的编码器。

II. 波形编码

在描述语音波形编码器之前，在有效的波形编码器设计中可以利用某些语音特性。这些特性包括波形振幅及功率的分布、语音谱的非平滑特性（及等效的自相关函数），音频语音的准周期性及信号中无声区域的存在。当通过这些叙述而深入时，我们会涉及到那样一些特性，而应用这些特性会要求增加编码器的记忆。例如，振幅与功率分布可以用来设计零记忆量化器。而除去语音多余度的编码器，正如语音频谱及周期性中表示的那样，可以要求 0.2～32 ms 的任一编码器记忆；利用语音波形的无声区域的 编 码系统可以要求秒级记忆或编码延迟。下面让我们扼要说明这些基本的特性。

A. 语音波形特征

语音波形的最基本特性也许是频带限制。频带限制于语音形成过程中开始，但另外的影响是典型的语音传输系统的有限带宽。例如，一般的音频电路带宽为 200～3200HZ。一般情况下，语音波形的有限带宽意味着可以用有限速率进行时间采样（低通信号的奈奎斯特速率是其中最高频率的 2 倍；商用电话一般的采样频率为 8000 HZ）。波形编码系统不仅以时间离散化为基础，而且也以不同方式的振幅离散化为基础。即使基本概念可以只用足够长的语音波形振幅的时间曲线（如图 3 所示）定性描述，但这些方式却利用了一系列的以形态上统计体制来定义的波形特性。

波形图直接表明语音的相对较强的准周期的浊音（用声带周期振动产生）与较低振幅的清音（由湍流气流的随机噪声通过某种压迫而产生）的显著不同。短时间信号统计反映出这两种音频声源的特点。

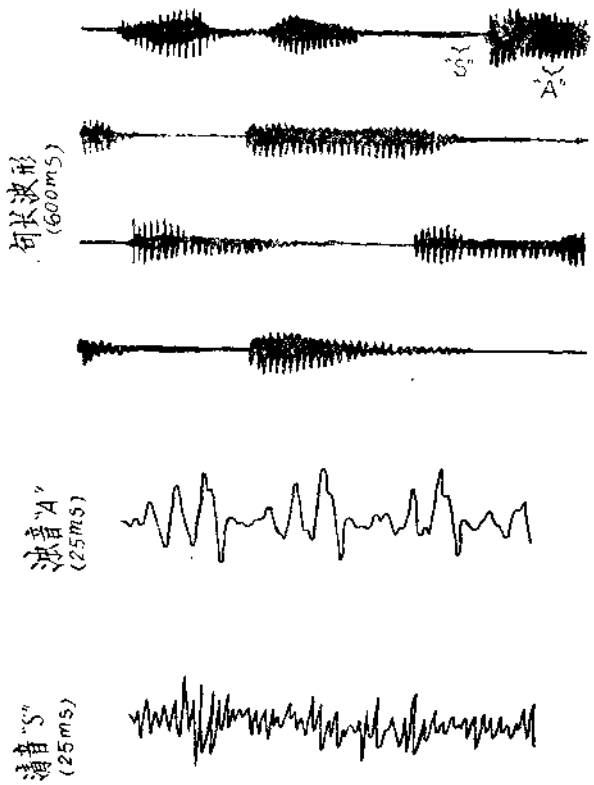


图 3　（句长）长时间和短时间(25ms)的语音时间波形

振幅分布

语音振幅的概率密度函数（PDF）通常用零或接近零振幅处有很大概率（与语音波形的中顿和低能部分对应）、极高振幅处有相当大的概率及这些极值之间振幅的单调下降函数来表征。对语音振幅的长时间平均概率密度函数的解析拟合可由高斯和拉普拉斯函数（双边指数）之和或 γ 函数表示。简单的高斯模型对短时间（20 ms 为准）的概率密度函数 PDF 通常是适合的。

自相关函数（ACF）

语音波形的振幅采样之间所具有的相关性用自相关函数 $C(n)$ 表示。其中相关量 c 是 t 时刻与 $t+n$ 时刻（零均值）振幅的乘积的（数学）期望（时间平均），归一化为振幅平方的期望（时间平均）值。根据定义，$C(n)$ 限制在 $[-1,1]$ 之间，且 $C(0)=1$。条件 $[C(n)=0;\ n\neq 0]$ 代表不相关（类似白噪声）振幅段。$C(1)$ 的正负值相对地表示慢变和快变振幅序列。

长时间（55秒）平均的 $C(n)$ 函数用图 4 所示的 8kHZ 语音例子说明。曲线上（面）组合及下（面）组合与低通滤波（0～3400HZ）和带通滤波（200～3400HZ）的语音相对应。在各个组合中上下极限相对于四个谈话者（两个男的和两个女的）的最大和最小值；其中中间的曲线给出了均值（四个谈话者进行平均）。对于 $n=1$，其平均值是 $C(1)\sim 0.9$，表示奈奎斯特采样(8kHZ)语音中很高的相邻采样的相关性。

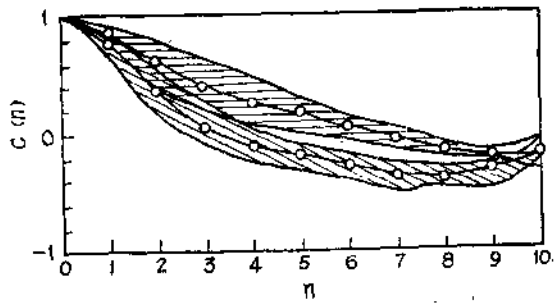


图 4　低通滤波语音(上面的曲线)和带通滤波语音（下面的曲线）的长时间自相关函数

短时间相关（根据平均 20ms 左右的观察）示于图 5 中。与上图不同，最大的 n 值（奈奎斯特采样中的自相关延迟）在这里是相当的大，它足以显示出第二个峯值。（该峯与经常存在于语音波形中的周期性有关）。事实上，图 5 也表示出短时间自相关函数 ACF 的从一个高周期、慢变化的语音部分(下面的)到一个非周期、变化较快的部分（上面的）的演变。

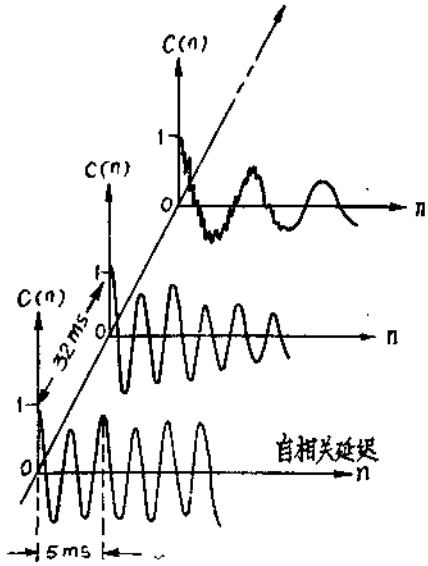


图 5 不同时间的语音波形的短时间自相关

功率谱密度(PSD)

语音中不同频率成分的平均概率用图 6 的长时间平均谱密度 $S(e^{j\omega})$ 曲线表示。很明显，波形的高频成分只占总的语音能量的极小部分。尽管如此，高频分量仍然是语音信息的很重要的成分，因此必须在编码系统中充分地表示出来。这对寻求图 6 所示的平均谱的解析模型是有用的。常用的模型由合成白噪声谱导出，它具有每倍频程衰减 6dB 的渐近线特性。该模型在离散时间系统中与一阶马尔柯夫过程相符。而且，还可用一指数衰减的 $C(n)$ 函数来表征。对平均语音谱的更为实用的模型是为每倍频程可能衰减 12dB 的特性（例如，二阶马尔柯夫过程），且在一适当频率上截断之。

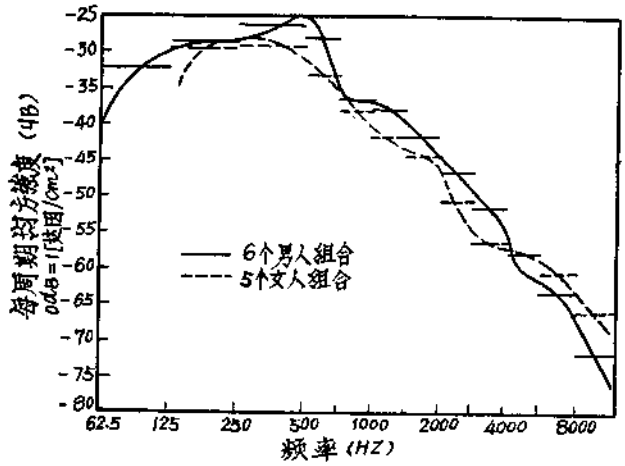


图 6 语音的长时间平均谱密度

正如相关性的描述那样，短时间统计特性(如在典型的 20ms 范围内的计量)也是有意义的。图 7 所示，短时间语音谱不总是简单的低通类型。清音波形可以具有高通谱；浊音虽然整个来看是低通，也显示出周部谐振特性，称为特征频率（formats）。事实上，特征频率的时间过程对语音的可懂度来说是至关重要的。

谱的平滑：予测或变换编码的增益

描述波形 $x(n)$ 的自相关函数 ACF 和功谱密度 PSD 可以等效成函数 $C(n)$ 及 $S(e^{j\omega})$ 的付立叶变换对：

$$S(e^{j\omega}) = |X(e^{j\omega})|^2 = \sum_{n=-\infty}^{\infty} C(n) e^{-jn\omega T} \qquad (1)$$

其中 $T=1/f_s$ 是采样间隔（秒）。为了实现时域（用 ACF 特性）或频域（用 PSD 特性）的任一码元的节省，可以利用语音波形的多余度。事实上，去掉多余度的编码器的最好的理论性能（所谓予测增益或有效变换编码增益）跟用确定的算术—几何平均之比来描述一样，与功谱密度 PSD 的不平滑性有关。如果用 $S_k[S(e^{j\omega k})$ 的缩写] 表示一定频率间隔上的 PSD 采样的话，其中 $K_{max}=N$ 相应于最高频率，且设 N 足够大，那么频谱平滑程度（SFM）用下式所示之比表示：

$$SFM=\left[\frac{1}{N}\sum_{k=1}^{N}S_k^2\right]\Big/\left[\prod_{k=1}^{N}S_k^2\right]^{\frac{1}{N}} \tag{2}$$

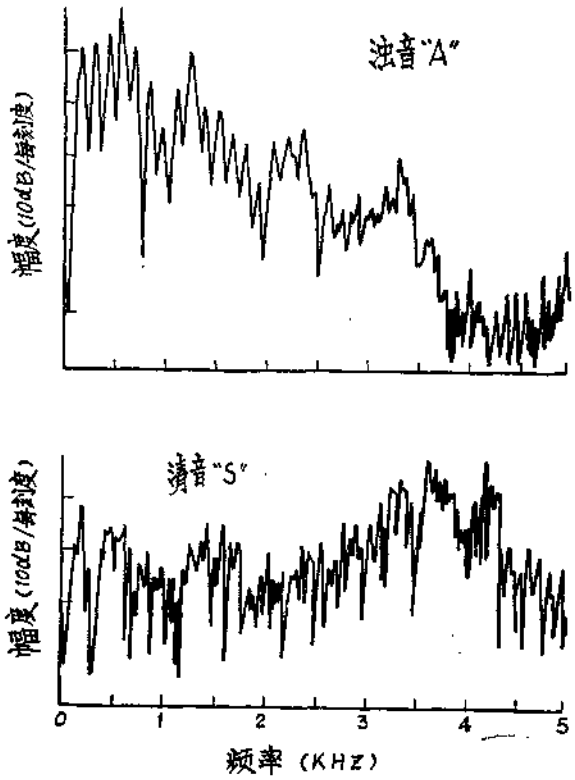


图 7 典型的浊音和清音的短时间谱

语音的 SFM 典型的长时间平均值的数量级为 8，其短时间值可以变化很大且落在 $2<SFM<500$ 范围内。

专用语音波形特性

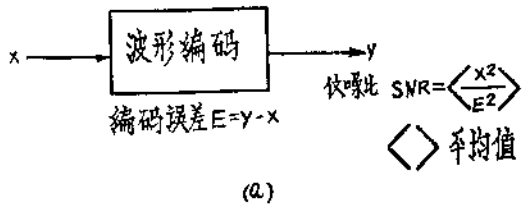
在我们的波形描述体系里最后是非常专门的语音信号特性。主要特性之一与音频语声的准周期性有关，另外的与信号的“能动因数”有关，换言之，与语音气流中寂静间隔有关。前者将在后面讨论自适应予测中适当描述，在评论语音的分组传输时，后者将按寂静统计及间隔变化来处理。

波形的保真标准：信噪比(SNR)

编码波形采样与原始输入波形之差习惯上定义为编码误差。在适当间隔范围内的平均值的平方叫做编码噪声。（在同一间隔内平均的）输入信号平方平均值与编码误差之比定义为信噪比（SNR）。该量常用分贝表示为 $10\log_{10}$SNR。在波形编码中总目的是在给定编码比特率 I 的情况下得到最理想的 SNR 最佳值。常常，但不总是，这就是最大可能的 SNR。

图 8 表示待量化幅度采样，最后直接或间接地表述成 B比特数码。当我们从简单的幅度量化器（如 PCM）讲到差分量化器(如DPCM 或予测编码)，直到分组量化器(如变换编码），B比特限制的意义就会明白了。图 8 中所用的均匀时间离散，不言而喻地也为本文的其余部分所采用。我们将语音看作一个频带限制的波形，因此可以利用采样理论。采样理论认为语音的频谱信息可以用采样频率为 f_s 的波形采样所保持，f_s 等于或

实际上稍高于奈奎斯特速率 f_{nyq}。后面我们将看到，对增量调制DM用过采样（$f_s>f_{nyq}$）能得到一定的收益。另方面，欠采样（$f_s<f_{nyq}$）会引起不容许的蜕化，即所谓折迭失真[余音1]。



(a)

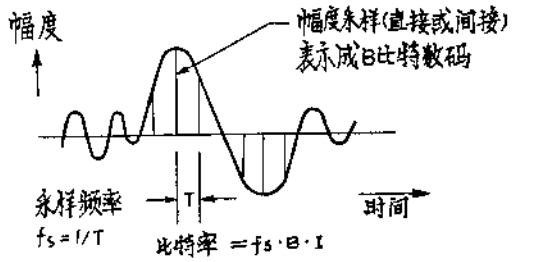


图 8：(a)波形编码器的 SNR 测量；(b)均匀采样及每样值B比特量化的表述。

我们可以看到，通常可将语音波形编码器的算法分成时域和频域两种。但更重要的是要强调不同类型的编码器能根据所利用的语音特性互相等效。例如，自适应予测 APC（时域方法）和自适应变换编码 ATC（频域方法）二者均利用语音信号的同样的多余度，而且它们的特性受同样的信息理论的上限所约束：具体地，即码率——失真函数（由高斯输入模型导出；如语音这样的非高斯信号，可以作得更好）

$$D(R)=2^{-2R}\cdot SFM \tag{3}$$

其中R为码率（比特/样值），SFM 是频谱平滑度(2)。

B. 时域编码

脉码调制（PCM）—通用量化器设计

波形编码器是利用将每一样值舍入到若干离散值中的一个的方法对幅度样值进行量化。在 B 比特量化器中，其离散的幅度电平数为 2^B。在量化理论中一个主要的结论是量化误差功率与量化阶距平方成正比；而且由于对一个给定的总幅度范围而言阶距与电平总数成反比，所以信号——量化误差比 SNR 可以定义为与 2^{2B} 成正比。如用对数作单位，则 SNR 随 B 线性增大。基本上是一个采样幅度量化器的 PCM 编码器可以有如下那样的性能公式

$$SNR_{PCM}(dB)=10\log_{10}SNR_{PCM}=(6B-\theta) \tag{4}$$

其中 θ 是与阶距有关的参量。

SNR 公式意味着，只要量化足够精细，量化误差样值就可以用附加噪声样值来作为模型。必须很好地记住，对粗糙量化（指 $B<5$）的误差波形有许多结构，且与输入语音本身非常有关，这些都应看作附加噪声。

要注意，在导出公式 (4) 时就假定，量化器的范围要用输入语音幅度来调节。如果语音幅度不超过量化器的过载点(在图 9a 表示的均匀 3bit 量化特性中为+4.0 和-4.0)，并且如果所有量化器范围都利用同样形式的话，这种要求对任何有意义概率都可以实现。实际上，这样的量化器对输入的调节是用下述两种技术之一来实现的，即非均匀量化或自适应量化。

非均匀量化(图9b)对经常出现的低振幅语音用较细微的量化阶来表示（因而有相对较小的噪声方差）；而用更粗糙的量化阶来照顾语音波形中偶然出现的大幅度偏移。语音振幅的平均分布是振幅的递减函数(图 9b 插图中注有"S"的虚线PDF)，非均匀量化则直接应用了这种语音特性。例如商业电话幅度量化器应用指数量化特性，它与对数压缩输入的均匀量化相等效。对数量化实际上比利用动态范围和空载噪声特性的 PDF 最佳量化(γ 最佳)更强。振幅压缩用在按所谓 μ 律或 A 律之一的对数量化中，对输入信号 $x(n)$ 而言两个特性均对 $x(n)=x=0$ 对称。$x>0$ 和 $x_{max}=1$ 的压缩信号 $x_c(n)$ 定义成：

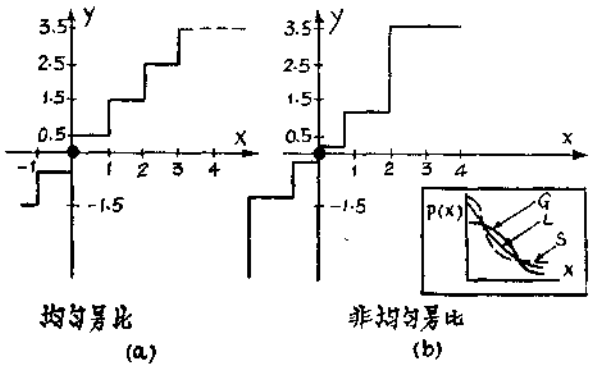


图 9 (a)均匀量化特性；(b)非均匀量化特性

$$\mu\text{律：}\ x_c=[\ln(1+\mu x)/\ln(1+\mu)];\quad \mu>0$$

$$A\text{律：}\ x_c=[Ax/1+\ln A],\quad 0\leqslant x\leqslant A^{-1}$$

$$=[1+\ln Ax/1+\ln A],\quad A^{-1}\leqslant x\leqslant 1 \qquad (6)$$

实际上或者叫折线型。一个 7bit (128 电平) 的语音对数量化器完全可以达到由 (4) 式所定的 35dB 的 SNR. 另方面，如图 9a 所示的均匀量化器，由于大振幅（虽然是少见的）偏移的缘故，需要约 12bit (4096 电平) 才能保证上述语音输入的信噪比。(录音 2)。

在典型的音频通信系统中，考虑到谈话者之间和谈话者本身的不同而具有的按音节平均的音量或功率的动态范围竟达 40dB。其时间不变性的非均匀量化已成为一个常见的解决动态范围难题的办法，即承认语音信号的大动态范围是由编码器输入的非平稳性或随时间变化过程而引入，便可得到较好的结果；所以真正最佳量化的办法是随时间变化或对输入信号自适应。

自适应量化是利用一个（均匀的或非均匀的）象手风琴那样随时间压缩扩展的量化特性。图10(a)(b)上两个瞬时的上述自适应量化图分别表示与低和高语音功率相适应。虽然，语音信号在一个长时间周期范围内有较大的动态范围，但输入功率电平变化缓慢，足以使遵循这些功率变化规律的简单适应方法的设计得到简化。例如，用单个字记忆的自适应量化中，在时间采样 $r-1$ 时的适应阶距被表示为

$$\Delta_{r+1}=\Delta_r\cdot M(\text{|量化输出}_r\text{|}) \qquad (7)$$

其中阶距乘数M仅是最后量化器输出的函数。由于适应跟踪的是量化器的输出而不是输入，虽然在这个方案中的阶距数

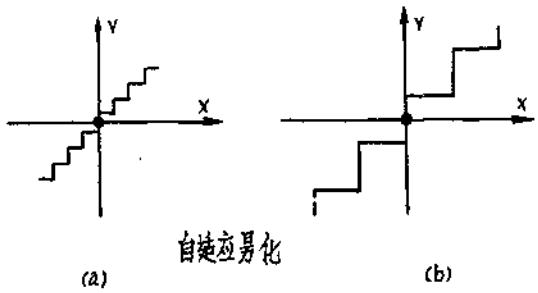


图 10 (a)小信号功率和(b)大信号功率的自适应量化特性

据无须明确的传输出去，在无误差传输情况下，仍然可以由收端准确地恢复。

差分脉码调制 (DPCM)

我们现在考虑一种比 PCM 更明智的利用语音中存在的多余度的编码技术。多余度或予测性可以用在时域或频域的任一种之中。还有一种多余度压缩码，它是对编码器的输出符号而不是输入语音进行处理。这些“无噪声”熵编码方案在一般编码文献中已提供了很好的论述。

语音波形的相邻振幅具有非常高的相关性（图 4）。这就说明语音振幅 $x(r)$ 及 $x(r-1)$ 之间的差值 D 的方差比 $x(r)$ 的方差小很多。这就直接提示一种方案，即不用波形振幅，而用波形振幅之差来描述语音。粗略地说，差分编码是取其差值进行量化，并对量化了的差值样值进行积分以恢复 $x(r)$ 的近似值。对给定的比特 B 而言，量化误差的方差与量化器输入的方差成正比。减少量化器输入方差一个因子 G，即能减少编码误差方差一个因子 G，同样地也就增加了 SNR 一个因子 G。如果输入信号相邻采样的相关系数 $C(1)$ 为 C_1（注意根据定义 $-1<C_1<1$），那么一阶差分编码的 G 可以表示成 $[2(1-C_1)]^{-1}$，这只有 $C_1>0.5$ 才有意义。更一般地，如果量化输入为 $x(r)-a_1x(r-1)$，那么可以看出 $a_1=c_1$ 时这个方差最小，且这种情况下予测增益为 $(1-C_1^2)^{-1}$，对 C_1 的所有取值，该增益均大于 1。形式上 $a_1x(r-1)$ 叫 $x(r)$ 的一阶予测；差分编码也叫予测编码。更一般的（线性）予测是 p 阶，其中输入采样 $x(r)$ 的估值是

$$\hat{x}(r)=\sum_{n=1}^{p}a_nx(r-n) \tag{8}$$

予测系数 a_n 的最佳值可根据 $x(n)$ 的自相关的前 p 个样值计算。

予测方式的译码包含一个积分过程，这可看作量化误差的相应积分或累加。量化噪声的累加由图 11 中反馈部分来控制，其中的予测要以量化值 $\hat{x}(r-n)$ 为依据：

$$\hat{x}(r)=\sum_{n=1}^{p}a_n\hat{x}(r-n)=\sum_{n=1}^{p}a_ny(r-n) \tag{9}$$

图 11a 代表一个通常所说的差分脉码调制（DPCM）的实际予测量化器。对任意的比特数 B，DPCM 的 SNR 均超过 PCM 的 SNR。这个增益除很粗糙的量化（如 $B=1$）外，很接近早先讨论的予测增益。p 阶 DPCM 的设计及性能在量化噪声忽略不计的假设（如 $B>2$）条件下可为

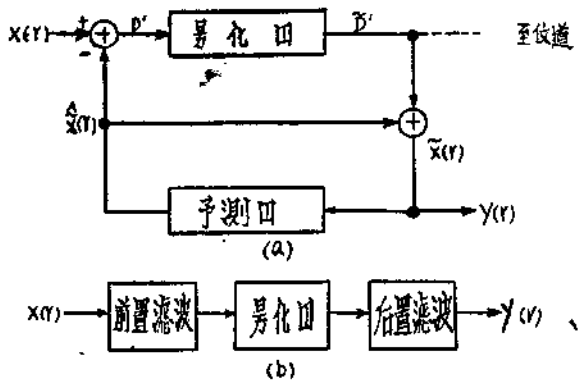


图 11 (a) DPCM 量化器和 (b)D*PCM 量化器

$$A_{oPt}=\Gamma^{-1}\Sigma \tag{10}$$

$$G_{oPt}=[1-A_{oPt}\cdot\Sigma]^{-1} \tag{11}$$

$$SNR_{DPCM}(dB)=SNR_{PCM}(dB)+10\log G \tag{12}$$

其中 A^T 是 p 阶予测矢量 $[a_1,a_2,\cdots a_p]$，G 是超过 PCM 的 SNR 增益（予测增益），Σ^T

是相关矢量 $[C_1, C_2 \cdots C_p]$，Γ 是相关矩阵 $[C_{ij}]$ $(i, j = 0, 1, 2 \cdots p-1)$。（如前所述，我们已将表示式简化成 $C_{ij} = C_{i+j} = C(|i-j|)$）。例如，用 8kHZ 采样和典型（长时间平均）的 $[C_0 = 1,\ C_1 = 0.85,\ C_2 = 0.562,\ C_3 = 0.308]$ 的相关系数，（非时变的）1～3 阶予测矢量分别为 $[0.85]$，$[1.13, -0.38]$ 和 $[1.10, -0.28, -0.08]$；而 16 和 24kHZ 采样的非时变一阶予测鉴于有相应较高的 C_1 值，则可采用典型的 0.95 和 0.98 的 a_1 值。

图 12 示出了 DPCM 作为予测阶数 p 的函数的增益 G（超过 PCM 的增益）的曲线，其输入分别是低通滤波和带通滤波的语音。这些曲线相应于图 4 的条件，在该情况下每个 G 特性曲线包括在四个谈话者范围内的最大值、最小值及平均值。可以看出，大体上予测阶数为 2 或 3 附近时予测增益趋于饱和。在时变或短时间 ACF 统计的自适应情况下，饱和点可推到 p 的最大值，例如图 5 那样。事实上，语音差分编码的最大成就取决于予测，包括频谱自适应（相邻采样予测，如用 $p \leqslant 12$）和基音自适应（相距较远的采样予测用 $20 \leqslant p \leqslant 120$）。这些将在下面的自适应予测编码（APC）一节进一步研究。

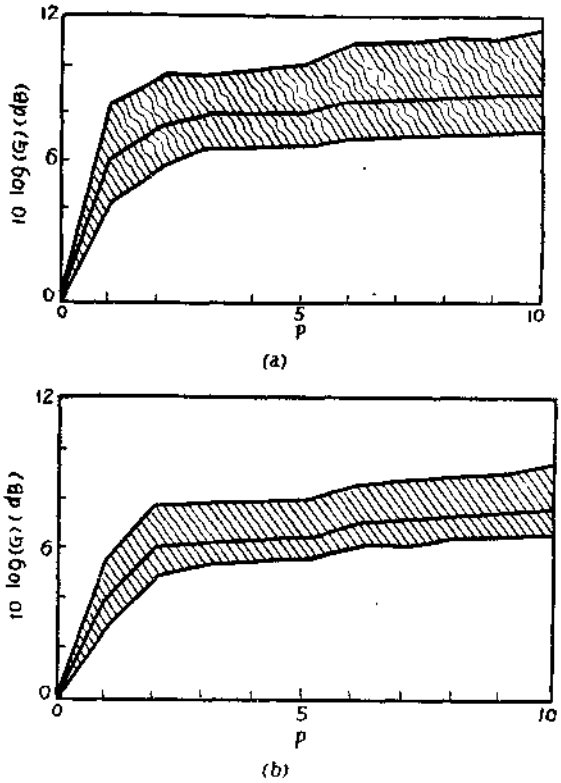


图 12 (a)低通滤波及(b)带通滤波语音的 DPCM 予测增益(dB)与予测阶数的关系曲线

参看图 11b，若对前置和后置滤波器正确设计，已译码语音中所形成的量化误差谱可以有快感的特性。这种电路叫新型 D*PCM。其均方误差特性介于 PCM 和 DPCM 之间。例如，如果前置滤波器限制为一阶予测，那么最佳予测系数为 $C_1(1 - \sqrt{1 - C_1^2})$（而不是 DPCM 的 C_1）；相对于 PCM 的增益为 $(1 - C_1^2)^{-\frac{1}{2}}$，而不是 DPCM 的 $(1 - C_1^2)^{-1}$。D*PCM 滤波器可以表示成一个部分变白滤波器（而 DPCM 予测试图全变白），后置滤波器是前置滤波器的倒置。虽然 DPCM 在上述讨论的意义上能完成 D*PCM 的任务，但如果希望具有某种频率加权的误差典型，有最小均方值的话，则后者更有吸引力。误差加权的一个例子是输入 PSD 的倒数—$S^{-1}(e^{j\omega})$。对语音而言，这种噪声加权可以得到比 DPCM 更好地抑制高频噪声的设计。

增量调制(DM)

DPCM 的一个特例是 1bit($B = 1$) 的形式——通常所说的增量调制（DM）。在这种 1bit（2 电平）的量化器里，予测误差被量化成仅有的两个值，即或正或负。如图 13 所示，其中 $y(r-1)$ 代表 $x(r)$ 的量化予测值。图 13a 代表 DM 的最简单形式，它是设想成

一个一阶予测的特殊情况下的理想积分器。

DM 编码器的工作原理很简单，在每个采样时刻 r，记下前一输入 $x(r)$ 与邻近它的后一阶梯 $y(r-1)$ 之差的极性 b_r，并按差值的方向不断改变阶梯函数。改变值 Δ_r 在非自适应（线性）增量调制 (LDM) 中是不随时间变化的，在自适应增量调制 (ADM) 中 Δ_r 随输入斜率变化。图13中画阴影部分表示产生斜率过载失真，它与 DM 用迅速变化的阶跃极性跟踪输入时出现的粒状噪声不同。在 $B=1$ 的特殊情况下，由于粗糙量化引入的这种不同是很重要的（不仅从统计意义上，而且从主观感觉上）。

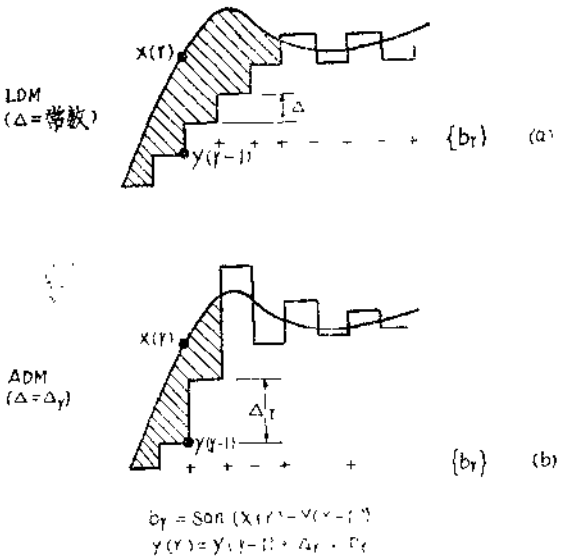


图 13 (a)线性增量调制(LDM)及
(b)自适应增量调制(ADM)

图 14 所示 DM 框图保持了 DPCM 的围绕量化器反馈的特点。(还指出 DM 可以利用有泄漏的积分器——一阶予测的一般形式，$a_1 \leqslant 1$，而不是图 13 的 $a_1=1$)。图中还表出 DM 中输入语音通常是过采样。很清楚，这是对 1bit 粗糙量化的补偿，即过采样可以增加相邻采样的相关性，增大予测增益，使量化器的工作更容易。容易看出，予测增益与 C_1 有关。在 DM 中，相关系数写作采样频率的函数 $C_1(f_s)$ 更合适。其结果是作为输入频谱许多重要特征，予测增益 $[1-C_1^2(f_s)]^{-1}$ 与 f_s^2 成正比。进而由于输入过采样的缘故，DM 编码中的量化噪声具有带外分量，仍然用输出低通滤波（如图 14 所示）以消除这些带外噪声分量。滤波增益典型地与 f_s 成正比，所以由此所得的 DM 的 SNR（包括予测增益及滤波增益的影响，与采样频率的立方成正比（也与 DM 中等于 f_s 的比特率成正比）。形式上对单积分（一阶予测）DM 编码有

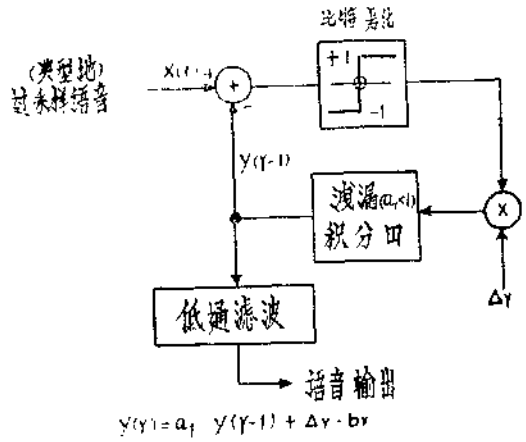


图 14 增量调制方框图

$$SNR_{DM}(dB)=10\log[k_1 f_s^3]=10\log k_1+30\log f_s \qquad (13)$$

其中 k_1 是一个与 1bit 量化性能和输入信号频谱形状(PSD)有关的常数。系统的最好特性显然由量化阶距 Δ 来实现，它要求设计成语音与带内噪声比为最佳值。若采用多比特

量化，实现最有用的量化加载的最好方法是让阶距自适应地跟踪输入斜率统计特性。在最简单的 1bit 记忆方法中，当 DM 出现斜率过载($b_r=b_{r-1}$)时可用因子 $P(>1)$ 扩展阶距 Δ_r，当出现抖动($b_r\neq b_{r-1}$)时用 $P^{-1}(<1)$ 来压缩阶距。最普通的适应规律是

$$\Delta_r=f(\Delta_{r-1},b_r,b_{r-1},b_{r-2}\cdots) \tag{14}$$

这种最普通算法的应用在后面传输误差中将进行讨论。

最后，一种设计成“双积分”的 DM 编码器可以得到以采样频率为单位的每倍频程 15 dB（而不是 9dB）的特性。这种编码器按其名称的意思，语音输入的编码则利用了它的二阶及一阶导数。与单积分比较，其优点随采样频率的降低而减少。进而在双积分设计（象一般高阶予测设计一样）中特别是涉及到某些自适应量化时稳定性的考虑是更重要的。

在即将结束 DPCM 和 DM 讨论之际可以看到，对所有的比特率 B 而言，多比特 DPCM 的 SNR 增益比 PCM 大 $10\log GdB$[见(12)式]，而 DM 也超过低比特率的 PCM，即（4）式的线性比特率函数比（13）式的对数比特率函数要差。典型地，当比特率为 32 kbit/s 或更低时，DPCM 及 DM 的 SNR 及主观感觉均优于 PCM [录音 3]。对 16 kbit/s 而言，DPCM 与 DM 的特性二者之间没有根本区别（录音 4）。同时，比特率等于或小于 16kbit/s 时，不同的编码器特性不可能用简单的 SNR 值充分地描述，而是与表征 DPCM 或 DM 的噪声结构的许多参量有关（录音 5），这与低比特率的 PCM 的情况是一样的。

最后，我们提一提那些与 DPCM 及 DM 具有同样基础的更复杂的波形表示原理。其中包括延迟（树）[Delayad(Tree)]编码、孔径(Aperture)编码及梯度搜索 (Gradient-Search) 编码。在这里由于篇幅所限，不能详细讨论。可以肯定地说，他们的许多优点取决于编码器中的较多的记忆及更广泛的处理能力。

自适应予测编码(APC)

予测编码系统讨论到这里还只限于用固定系数的线性予测器。然而，由于语音信号的非平稳性，固定予测器不可能总是有效地予测信号值。于是予测器必须随时变化，以适应语音信号频谱包络的变化及有声语音周期的变化。

语音自适应予测做成两个分离级最方便：即一级利用连续语音采样之间的相关性（或等效成语音信号的短时间频谱包络的不均匀性）予测；另一级利用有声语音的准周期性（或表征有声基音谐波线频谱的精细结构）予测。

根据频谱精细结构的予测

予测周期性信号即时值的简单方法是用前一个或前几个周期的信号值与其相等。对于语音，予测器已提供一些增益的调整及算出某一周期与另一周期的振幅变化。予测器可以用 z 变换表示成

$$P_d(Z)=\beta Z^{-M} \tag{15}$$

其中 M 代表 2～20ms 范围内的较长延迟，β 是比例因子。大多数情况下，延迟相应于基音周期（或可能为基音周期的整数倍）。予测量由相邻基音周期之间的相关性决定。这样的相关性典型地随语声及谈话者不同而显著地变化。对浊音语音，平均予测增益约 13 dB。对清音语音，其本质上相当于噪声，因子 β 很小，M 值是无关紧要的。

确定(15)式的未知参数 β 和 M 是比较直观的。延迟 M 的选择要做到相隔 M 个采样延迟的语音采样之间的相关性最大。参数 β 为

$$\beta = \langle S_n S_n - M \rangle_{av} / \langle (S^2)_n - M \rangle_{av} \tag{16}$$

其中 S_n 为第 n 个语音样值，$\langle\ \rangle_{av}$ 表示给定时间间隔内所有样值的平均，而在这个间隔范围内予测器为最佳。

在予测即时采样时的基音延迟两边额外增加采样，予测增益就可以提高。予测器用 z 变换表示为

$$P_d(z) = \beta_1 z^{-M+1} + \beta_2 z^{-M} + \beta_3 z^{-M-1} \tag{17}$$

这三个振幅系数提供了一个由频率而定的增益因子，这是很需要的。因为在语音的较高频率上有声音的周期表现并不很强，语音波形的采样也能造成较高频上相邻基音周期之间相关性减弱。在计算机模拟试验中，3 阶基音子测器的平均予测增益比一阶增加近 3dB。

根据短时间频谱包络的予测

短时间语音频谱包络由声域的频率响应决定，对浊音语音还由声带音色脉冲频谱决定。予测器特性的 z 变换表示式为

$$P_s(z) = \sum_{k=1}^{p} a_k z^{-k} \tag{18}$$

其中 z^{-1} 表示一个采样间隔延迟，$a_1, a_2 \cdots a_p$ 是 p 个予测系数。用 8KHZ 采样的语音的 p 的典型值是 10。予测系数 a_k 由持续时间连续变化 10～30ms（在该时间间隔内音域结构可以近似看作是平稳的）的自适应方式中予测剩余功率最小而定。对确定随时变化的予测系数的方法，除协方差矩阵周期地（每 10～30ms 变化一次）不断变化而外，与 DPCM 中所用固定予测方法之一完全相同。

随着 p 的增大在极限情况下予测渐近值唯一地由短时间信号频谱包络决定，具体地由(2)式的平滑度 SFM 决定。平均予测增益(dB)随予测系数的数目而变化，对 10KHZ 采样的浊音语音情况下该变化如图 15 所示。当 $p=10$ 时，浊音语音的平均予测增益为 14dB。千万要记住，在予测编码中，输入信号采样是由在收端所得的前面的复现信号采样来予测，而不是由实际的输入信号前面的采样所予测。因而输出信号含有噪声，量化粗糙时予测增益要下降。

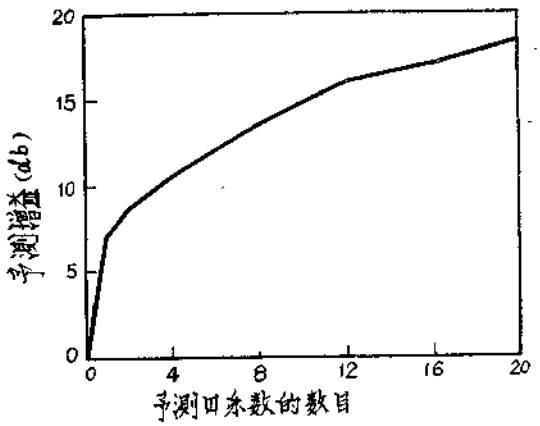


图 15　10KHZ 采样时浊音语音的予测增益(dB)随予测系数数目的变化

两种自适应予测的组合

当两种予测以任意次序串接时，所得的予测算子 $[1-P(z)]$ 实际是 $[1-P_d(z)]$ 与 $[1-P_s(z)]$ 的乘积。予测 $P(z)$ 能使予测增益比每个 P_d 或 P_s 单独作用时的要高。但是，总的予测增益(dB)不是 P_d 和

P_z单独作用于语音信号的增益和。组合应用时，第一个予测器提供大部分增益(典型为13-14dB)，第二个予测器（它现在所处理的信号的可予测性比原始信号要低）提供其余的部分（典型为其余的3dB)。

最佳予测编码器的感觉标准

到现在为止，我们的重点是放在最小量化噪声功率上。但为了确保语音信号失真感觉最小，这就需要研究量化噪声的频谱及其与语音频谱的关系。听觉掩蔽理论假设在语音能量集中的频率范围(如特征频率范围)内噪声能部分地或全部被语音信号掩蔽。于是编码器中可觉察的噪声大部分来自信号电平低的那些频率范围，而且，所需要最小化的也许不是量化噪声功率，而是它的响度。量化噪声响度最小化的方法已超出本文范围。但我们将提出改进予测编码中达到感觉上低响度的量化噪声频谱的简单方法。图16示出了一种能够得到任意要求的量化噪声频谱的予测编码器。在图16的反馈编码器中，输出噪声与量化噪声频谱之比为

$$\lambda(\omega)=\left|\frac{1-F(e^{j\omega T})}{1-P(e^{j\omega T})}\right|^2 \tag{19}$$

其中T为采样间隔。(19)式表明$\log\lambda(\omega)$平均值的主要限制是

$$\frac{1}{\pi f_s}\int_0^{\pi f_s}\log\lambda(\omega)d\omega=0 \tag{20}$$

这里的f_s是采样频率。表示在dB坐标上$\log\lambda(\omega)$的平均值为0dB．然而滤波器F将噪声功率从一个频率重新分配到另一个频率。于是某一频率噪声的减弱只能是以另一频率噪声的增加为代价。由于编码器产生的大部分可觉察噪声来自低电平信号部分，所以滤波器F可以作成在这个范围内减弱却在特征频率范围内增加。而在特征频率范围内噪声能有效地为信号所掩蔽。

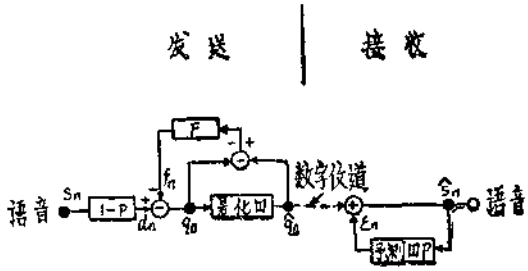


图 16. 可以得到任意要求的量化噪声谱的通用予测编码器框图

几种任意的但例证性的反馈滤波器F的选择应考虑：(i)$F=P$：这将导致最小的非加权噪声功率。输出噪声频谱是白噪声谱。在特征频率上达到很高的SNR、但在特征频率之间则很差（这时信号频谱的幅值很低)。(ii)$F=0$：这意味着没有反馈；输出噪声频谱与原信号一样但向下移动了，如果我们听觉对所有频率的量化失真一样灵敏的话，这是一种很好的选择。(iii)一个中间的设计例子为

$$F=P(\alpha z^{-1}) \tag{21}$$

这里α是为了增加$1-F$零点的带宽而附加的系数。增加频带使得在信号低电平范围内噪声减少的同时在特征频率范围内的噪声达到极值。带宽增加400HZ就能得到满意的

结果。图 17 示出一个与相应的语音频谱一起的输出噪声频谱包络的实例。

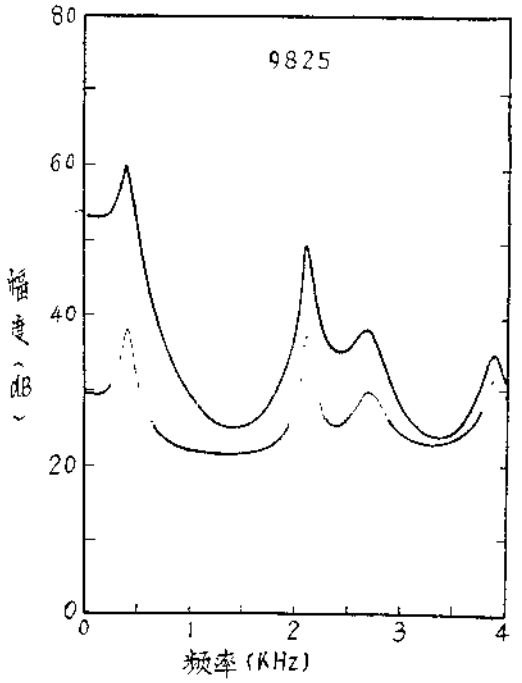


图 17　表示为了减小感觉失真[F为(21)式]所得到的输出噪声谱包络（虚线）及相应的语音谱(实线)的例子

通用语音予测编码器

语音信号的通用予测编码器框图如图18所示。语音信号的高频部分在分析之前用传递函数为 $[1-0.4z^{-1}]$ 的滤波器予加重。频谱包络予测器是 P_s，第二予测器 P_d 按基音周期予测。为了得到样值 f_n，对量化噪声进行滤波且限制其峯值。复合信号 $q_n=d_n-d'_n-f_n$ 用自适应三电平量化器等效。量化器的参数选择为高斯 PDF 最佳。(对均匀间隔三电平量化器，量化器输出电平之间的最佳间隔为予测误差有效值的 1.22 倍)。予测器和量化器参数每 10ms 调整一次。适当选择 F 可得到近 20dB的平均 SNR（比 $F=P$ 低，比 $F=0$ 高），且输出语音质量主观上与具有 33dB SNR 的 7bit 对数（压扩）PCM 的质量接近。与予测编码一起的非均匀噪声频谱能得到约 13dB 的主观 SNR 增益。

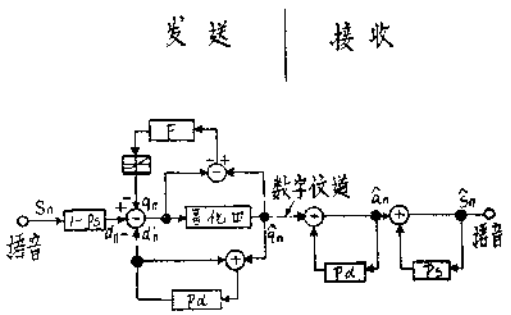


图 18　用两级予测的通用予测编码器框图：P_s—按短时间谱包络予测；P_d—按基音周期予测。滤波器F选择使量化噪声的感觉失真减小

在录音（录音 6）中包括用三电平及二电平量化器量化的两个带有量化误差的语音例子。在这些录音中予测参数未被量化。