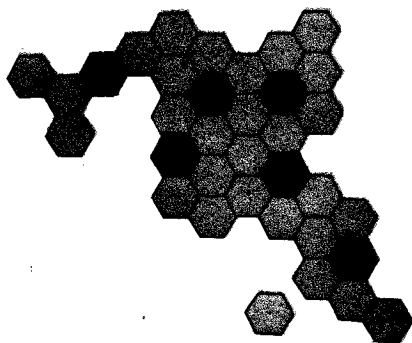




矢量量化原理 与应用

11.76

内 容 简 介

本书介绍矢量量化的基本原理,各种矢量量化系统的构造和设计方法,以及矢量量化在语音编码、图像编码、语声识别与合成中的应用,并简单介绍了用硬件实现矢量量化器的方法。

本书可供从事通信工程,信息工程,计算机通信,语音、数据和图像信号处理专业的科技人员与教学人员阅读和参考,并可作为上述专业研究生课的教材。

矢 量 量 化 原 理 与 应 用

胡 征 杨有为 编著

责任编辑 叶德福

西安电子科技大学出版社出版发行

西安电子科技大学印刷厂印刷

新华书店经销

开本 850×1168 1/32 印张 5 24/32 字数 138 千字

1988 年 6 月第 1 版 1988 年 6 月第 1 次印刷 印数 1-2 000

ISBN 7-5606-0044-1/TN·0016 定价: 1.00 元

前 言

随着计算机和大规模集成电路的飞速发展和微处理机的广泛应用,数字信号的分析与处理技术得到了很大的发展,并已广泛地用于通信、雷达、自动化等领域中。数字信号明显的优点是便于传输、存贮、交换、加密和处理等。

模拟信号变换为数字信号可分两个步骤:第一步是将模拟信号在时间上离散化,抽样定理解决了这个问题;第二步是将连续取值的信号幅度离散化,即用离散取值的信号去逼近连续取值的信号,这就是量化。

量化可以分成两大类:一类是标量量化,另一类是矢量量化。标量量化是把抽样后的信号值逐个进行量化,而矢量量化是先将 K 个($K \geq 2$) 抽样值形成 K 维空间 R^K 中的一个矢量,然后将此矢量一次进行量化。它可以大大地降低数码率,因此很受人们的重视。

矢量量化总是优于标量量化的。这是因为矢量量化有效地应用了矢量中各分量间的四种相互关联的性质(线性依赖性、非线性依赖性、概率函数的形状以及矢量维数)来消除多余度。而标量量化只利用线性依赖性和概率密度函数的形状来消除多余度。从理论上讲,正如山农的率-失真理论指出:当编码长度 $K \rightarrow \infty$ 时,速率与失真的关系可以达到率失真界。一般地说, K 愈大,此关系离率失真界愈近。因此,矢量量化优于标量量化,这也是率-失真理论的结论。矢量量化只不过是率-失真理论的一种具体实现途径。

矢量量化的应用范围是很广泛的,它既可以应用于语音编码,语声识别与合成,也可以应用于图像数据的压缩。可以说,凡是

要量化的地方都可以应用矢量量化。因此,开展这一领域的研究是非常必要的。

编写本书的目的在于使读者对矢量量化的原理和应用有一个基本的概念,缩短读者对矢量量化这一课题的了解和熟悉过程,从而进一步推动矢量量化课题的研究工作。

在叙述方法上,我们力求讲清物理概念,不过多地强求数学上的严密性。对各种矢量量化系统,没有打算作更详尽的描述。有些新出现的量化方法还没有涉及到,这是因为这些内容或者涉及较深的数学问题,需要较多的数学准备知识;或者这一研究课题正处于发展阶段,现在还很难作深入的介绍。

本书是在《矢量量化原理与应用》讲义的基础上重新编写的。这次编写时,总结了前一阶段为研究生讲授这一课题的经验,并根据我们在矢量量化课题研究*中的新进展和某些成果,增加了不少新内容。在我们的课题研究工作中,寇卫东博士做了大量的工作,并与本书编者以及刘天民,叶雷等五人共同编写了原讲义。我们对他们所作的贡献,表示感谢。在本书编写过程中,得到西北电讯工程学院出版社的大力支持和鼓励,黄饒卿为我们绘制了全书的草图,我们在此表示衷心的感谢。由于水平有限,书中难免有错误和不妥之处,请读者批评指出。

编著者

1987年7月于西北电讯工程学院

*本课题为国家自然科学基金资助的课题

目 录

第一章 概述

一、引言	1
二、无记忆标量量化	6
1. 均匀量化	6
2. 压缩与扩张	10
3. 最小二乘方量化与 Lloyd-Max 算法	15
三、有记忆标量量化	24
1. 差值脉码调制(DPCM)	25
2. 增量调制(ΔM)	28

第二章 矢量量化的基本原理

一、矢量量化的定义与矢量量化器的构造	30
1. 矢量量化的定义	30
2. 失真测度	33
3. 矢量量化器	34
二、最佳矢量量化器	37
1. 最佳矢量量化器概念	37
2. 设计最佳矢量量化器的必要条件	37
三、矢量量化器的设计算法	39
1. LBG 算法	39
2. 初始码书的选定与空胞腔的处理	43
3. 其它设计算法	45
四、矢量量化压缩数据的机理	47
五、矢量量化与标量量化的比较	54

第三章 矢量量化系统

一、基本矢量量化系统	56
1. 矢量量化器的复杂度	57
2. 时间复杂度	57
3. 空间复杂度	59
二、树搜索矢量量化器	60
1. 树搜索原理	60
2. 树结构的设计	62
3. 树搜索矢量量化器的复杂度	64
三、多级矢量量化器	70
1. 多级矢量量化器的构造	70
2. 多级矢量量化器码书的设计	71
3. 两级矢量量化器	72
4. 多级矢量量化器的复杂度	74
四、乘积码矢量量化器	76
1. 增益/波形矢量量化器	76
2. 分离均值矢量量化器	80
五、格形矢量量化器	81
1. 根格的定义及其某些性质	82
2. 输入矢量在 K 维整数格 Z^K 中的最近格点	84
3. 输入矢量 X 至 E_0 中的最近格点	86
4. 格点的标号	87
5. 根格的格点	90
六、变换域中的矢量量化	91
第四章 有记忆矢量量化器	
一、有记忆矢量量化器的分类	94
二、反馈矢量量化器	95
1. 预测矢量量化器(PVQ)	95
2. 有限状态矢量量化器(FSVQ)	107

三、自适应矢量量化器	112
1. 增益自适应矢量量化器	113
2. 自适应矢量预测编码(AVPC)	115
第五章 快速搜索算法	
一、搜索算法复杂度的度量方法	130
1. 三元组度量法	130
2. 乘法次数度量法	130
二、投影法	131
三、超立方体法	136
四、最小最大法	139
第六章 矢量量化的应用	
一、矢量量化在图像编码中的应用	144
1. 设计任务要求	145
2. 方案选择	145
3. 设计步骤	147
4. 实验结果	149
二、矢量量化在语音编码中的应用	151
1. VQ LPC 声码器	151
2. 语音波形编码器的硬件实现	157
三、矢量量化在模式识别中的应用	163
1. 采用类音素模板的单词识别系统	163
2. 隐马尔可夫模型单词识别系统	166
参考文献	

第一章 概 述

一、引 言

数字通信系统和模拟通信系统相比较,前者具有抗干扰能力强,保密性好,可靠性高,便于传输、存贮、交换和处理,传输中的差错可以设法控制,复用方便,设备便于生产以及易于集成化等优点;另外,它便于实现综合业务数字网(ISDN)。

把模拟信号(如语音,图像信号等)变为不仅在时间上是离散的而且在幅度上也是离散取值的数字信号,通常可以分两步完成。第一步是抽样,以得到时间上离散化的信号;第二步是量化,以得到幅度取值上离散化的信号。一个模拟信号 s , 经过预滤波器(最高截止频率为 f_H)后,变成一个频带限制在 $0 \sim f_H$ 内的时间连续信号 x_t 。根据奈奎斯特抽样定理,如果以不小于 $2f_H$ 次/s 的速率对 x_t 进行等间隔抽样,则信号 x_t 将被抽样值 x_{t_n} ($n = 1, 2, \dots$) 完全确定,即

$$x_t = \sum_{n=-\infty}^{\infty} x_{t_n} \frac{\sin 2\pi f_H(t - n\Delta t)}{2\pi f_H(t - n\Delta t)} \quad (1-1)$$

式中 $t_n = \frac{n}{2f_H} = n\Delta t$,

x_{t_n} ——第 n 次抽样瞬时幅值(样值);

$2f_H$ ——抽样频率;

Δt ——奈奎斯特抽样周期。

实际上,考虑到滤波器等电路特性,抽样频率通常选为信号

最高频率的 2.5 倍左右。例如，对频带宽度为 300 Hz—3400 Hz 的电话信号，抽样频率常选为 8 kHz。

模拟信号经抽样后，样值序列 $\{x_n\}$ 是一个离散的时间随机序列。为了传输、存贮和处理这个样值序列，必须将 $\{x_n\}$ 量化为一个只有有限个幅度值的集合 $\{\hat{x}_n\}$ ，以便可以用一个有限符号集合 $\{c_n\}$ 表示之，这便是编码。有时将量化和编码统称为信号的编码。以上这个过程就是模拟/数字 (A/D) 变换过程，如图 1-1(a) 所示。序列 $\{c_n\}$ 经信道传送到译码端，由于引入误码，译码器输入信号为序列 $\{c'_n\}$ ，如图 1-1(b) 所示。序列 $\{c'_n\}$ 经译码器译码后，得到量化后的样值序列 $\{\hat{x}'_n\}$ ，再将它通过低通滤波器，便得到原模拟信号 x_t 的恢复信号 x'_t ，这就是 D/A 变换过程。

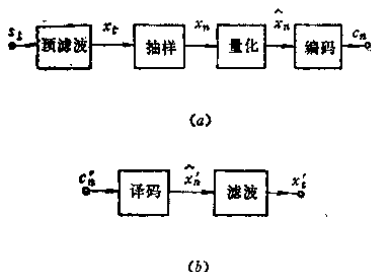


图 1-1 A/D 和 D/A 变换方框图

模拟信号经抽样后，因为其幅度的取值仍可以有无穷多个，所以它的幅度还是一个连续变化的量。量化就是将抽样值的幅度，按一定规律进行“分级”或“分层”。也就是把整个抽样值的幅度范围，划分成一组有限个幅度区间 $\Delta x_n (n = 1, 2, \dots)$ 。通常把区间 Δx_n (如图 1-2 中的 $\Delta x_1, \Delta x_2, \dots$) 称为量化区间。对落入某

个区间内的所有抽样值，赋予相同的幅度值 $\hat{x}_n$ (如图 1-2 中的 $\hat{x}_1, \hat{x}_2, \dots$)，我们把 $\hat{x}_n$ 称为量化值。

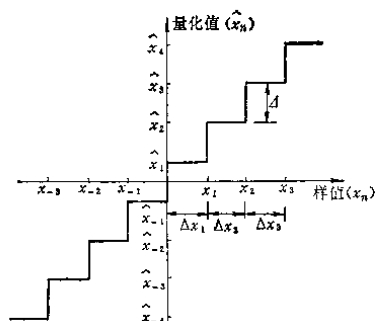


图 1-2 量化特性

由图 1-2 可见，对于处在量化区间 Δx_1 内的所有样值，对应量化器的输出均为 $\hat{x}_1$ 。同样，在量化区间 Δx_2 内的所有样值，对应量化器的输出均为 $\hat{x}_2$ 。对于大于抽样值 x_3 的所有样值，量化器输出均等于量化值 $\hat{x}_4$ 。量化就是这样把一个样值序列 $\{x_n\}$ ，变为一个取值范围有限的而且幅度值是离散的量化值序列 $\{\hat{x}_n\}$ 。图 1-2 中的 Δ 称为量化阶距。上述的量化是以一个一个单独抽样值 $x_n (n=1, 2, \dots)$ 为基础，通过合理划分量化区间 $\Delta x_n (n=1, 2, \dots)$ 和确定量化值 $\hat{x}_n (n=1, 2, \dots)$ ，来达到失真最小或量化噪声最小。通常把这种量化称为标量量化 (Scalar Quantization)。

理论上，信号量化的中心问题是：一方面要求离散取值的幅度尽可能地逼近原来连续取值的幅度，即失真或量化噪声要越小越好；另一方面又要求编码时所用的码位数（即比特数）越少越好。

这两者是矛盾的，山农的率-失真理论(Rate-Distortion Theory)解决了这个矛盾。当然，在实用方面，存在一个复杂度(Complexity)问题，这将在第三章中详细讨论。

人们最早认识的话音编码是脉码调制(PCM)，现在仍在大量使用。其速率需要 64 kbit/s(千比特/秒)，才能得到令人满意的话音质量。由于它需用的速率较高，人们就不断地研究出各种各样压缩数码的编码方法，如差值脉码调制(DPCM)，增量调制(AM)，自适应差值脉码调制(ADPCM)，子带编码(SBC)，变换域编码(ATC)以及声码器(Vocoder)等。上述各种方法都是基于标量量化技术。自 70 年代末，矢量量化(Vector Quantization)技术得到了发展。所谓矢量量化，就是先将 K 个($K \geq 2$)抽样值形成 K 维空间 R^K 中的一个矢量，然后将此矢量一次进行量化，并设法使其失真或量化噪声最小。这种量化方法可以大大降低编码速率，因此很受人们的重视。

矢量量化的发展，大致可以分为两个阶段：第一阶段约从 1956 年起到 1977 年为止。1956 年 Steinhaus 第一次系统地阐述了最佳矢量量化问题。1957 年在 Lloyd 的“PCM 中的最小平方量化”一文中给出了如何划分量化区间和如何求量化值问题的结论。约与此同时，Max 也得出了同样的结果。虽然他们讨论的都是标量量化问题，但他们的算法对后来的矢量量化的发展有着深刻的影响。1964 年 Newman 研究了正六边形定理。1977 年 Berger 的《率失真理论》一书出版。总起来说，这一阶段的工作多是理论性的、初步的和探索性的，但它们为第二阶段的发展奠定了一定的基础。

第二阶段约从 1978 年到现在。1978 年，Buzo 第一个提出实际的矢量量化器。他的量化系统的组成为两步：第一步将语声作线性预测分析，求出预测系数；第二步对这些系数作矢量量化。于是得到压缩数码的语声编码器。1980 年，Linde, Buzo 和 Gray

将 Lloyd-Max 算法推广,发表了第一个矢量量化器的设计算法,通常称为 LBG 算法。这就使矢量量化的研究向前推进了一大步。在这一时期,人们对矢量量化各个方面的问题展开了全面的研究。其中主要的是对失真测度的探讨,码书的设计,各种矢量量化系统的研究,快速搜索算法的寻找等。不过,大多数工作仅止于计算机模拟,而硬件实现还是近二、三年的事。

矢量量化研究的进展是很快的,1980年美国加州信号公司的 Wong 和 Juang 等人在原来编码速率为 2.4 kbit/s 的线性预测声码器上,仅将滤波系数由标量量化改为矢量量化,就可使编码速率降低到 800 bit/s,而声音质量基本未下降。1983年,美国 BBN 公司的 Makhoul 等人研制了一种分段式声码器。由于该声码器采用了矢量量化,所以可以用 150 bit/s 的速率来传送易懂的话音,这使人们感到震惊。最近二、三年,则已向深度发展。在已经提出的各种矢量量化方法和系统的基础上,再与其它编码技术相结合,得到了更好的系统构造方法。在图像数据的压缩和语音识别的应用方面,矢量量化研究也得到很快的发展,提出了各种各样的矢量量化系统,用硬件实现矢量量化系统的方法也日益增多。因此,目前已处于探讨矢量量化的实际应用阶段。

矢量量化的理论基础是山农的率-失真理论。率-失真理论是对给定的失真 D , 我们可以计算率-失真函数 $R(D)$, $R(D)$ 定义为: 在给定的失真 D 条件下, 所能够达到的最小速率(用每维计算); 或者反过来, 可以计算率-失真函数的逆函数 $D(R)$, 称 $D(R)$ 为失真-率函数, 它定义为: 在给定的速率(以 bit/s 计算)条件下, 所能够达到的最小失真。 $D(R)$ 或 $R(D)$ 所给出的编码工作性能极限, 不仅适用于矢量量化, 而且适用于所有信源编码方法。 $D(R)$ 是在维数 $K \rightarrow \infty$ 时 $D_K(R)$ 的极限, 即

$$D(R) = \lim_{K \rightarrow \infty} D_K(R)$$

率-失真理论指出, 利用矢量量化, 编码性能有可能任意接近率-失真函数, 其方法是增加维数 K 。率-失真理论在编码实践上有重要指导意义: 在编码中, 我们可以将实际的方法与率-失真函数相比较, 看它还有多少潜力可挖。总之, 率-失真理论启示我们, 矢量量化具有很大的优越性。不过, 率-失真理论是一个存在性定理, 并非是构造性定理, 因为它并未给出如何构造矢量量化器的方法。

矢量量化总是优于标量量化的, 这是因为矢量量化能有效地应用矢量中各分量间的 4 种相互关联的性质来消除数据中的多余度。这 4 种相互关联的性质是线性依赖(相关性), 非线性依赖(统计不独立), 概率密度函数的形状以及矢量量化的维数。而标量量化仅能利用线性依赖和概率密度函数的形状来消除多余度。但是, 标量量化与矢量量化有密切的关系, 矢量量化的一些算法与定理都是由标量量化的算法与定理推广而得到的。下面我们先讨论标量量化的一些重要概念, 这对研究矢量量化是很有益的。

二、无记忆标量量化

在无记忆量化中, 量化值 $\hat{x}_n$ ($n=1, 2, \dots$) 只与当前输入样值 x_n 有关。由于量化区间 Δx_n ($n=1, 2, \dots$) 的划分和量化值 $\hat{x}_n$ 的选取方法不同, 可以得到不同的量化方式。下面分别讨论几种不同的量化方式, 以便为研究矢量量化作些准备。

1. 均匀量化

均匀量化是量化区间 Δx_n ($n=1, 2, \dots$) 和量化阶距 Δ_n 均相等的量化, 即

$$\begin{aligned}\Delta x_1 &= \Delta x_2 = \Delta x_3 = \dots = \Delta \\ \Delta_1 &= \Delta_2 = \Delta_3 = \dots = \Delta\end{aligned}$$

图1-3(a)给出的量化特性,是原点出现在阶梯形函数上升部
分的中点。具有这种量化特性的量化器,称为“上升中点”(Mid-rises)型量化器。图1-3(b)给出的是原点出现在阶梯形函数水平部分中点的情况。具有这种量化特性的量化器称为“水平中点”(Mid-tread)型量化器。

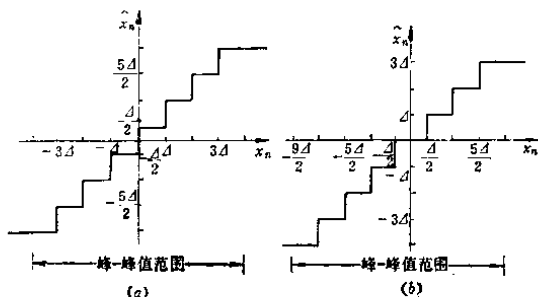


图 1-3 两种常用的均匀量化器特性

由图 1-3 可见,“上升中点”型量化器的输入/输出特性的特点是正负量化值(即量化电平)的数目相同,也就是说,这些量化值对原点对称的。而“水平中点”型量化器的输入/输出特性的特点是负的量化值数目比正的量化值数目多一个。然而后者有一个零量化值,前者却没有零量化值。由此可见,均匀量化主要有两个参数:

- (a) 量化值的数目(即量化电平或分层数目);
- (b) 量化阶距 Δ 的大小。

量化值数目 $N = 2^n$ (n 是二进制数字的位数,即编码比特数)。若 $n = 3$ (三位二进制码),则 $N = 8$,表示有 8 个量化值。通常设计量化器时,主要工作是选取合适的量化值的数目 N 与量化阶距

Δ 的大小, 以覆盖输入样值的幅度范围。很明显, N 选得大一点, Δ 选得小一点, 量化的性能就好, 但编码速率就高。

信号量化噪声比 $(S/N_q)_{dB}$ 是用来描述量化器性能的一个重要指标。由图 1-4 可见, 在量化阶距 Δ 相等的情况下, 落在 $[\hat{x}_K - \Delta/2]$ 与 $[\hat{x}_K + \Delta/2]$ 范围内的所有样值 x_n , 量化后均等于量化值 $\hat{x}_K$ 。同样, 落在 $[\hat{x}_{K+1} - \Delta/2]$ 与 $[\hat{x}_{K+1} + \Delta/2]$ 范围内的所有样值 x_n , 量化后均等于量化值 $\hat{x}_{K+1}$ 。因此, 量化后的信号 $\hat{x}_n$ 与原信号 x_t 之间存在着一定误差, 这种误差称为量化噪声。

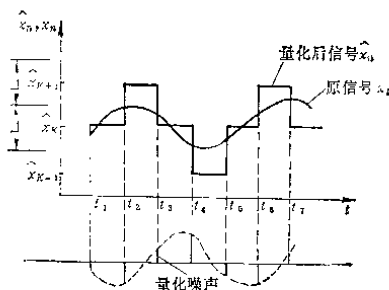


图 1-4 均匀量化的量化噪声

在任何一个量化区间 Δx_n 内, 产生的误差为 $(x - \hat{x}_K)$, 若话音信号 x 的概率密度用拉普拉斯函数来近似表达

$$P(x) = \frac{1}{\sigma_e \sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma_e} |x|} \quad (1-2)$$

式中 σ_e 是话音信号的有效值, 那么落在量化区间 $(\hat{x}_K - \Delta/2, \hat{x}_K + \Delta/2)$ 内的量化噪声功率(即误差的均方值)为

$$N_P = \overline{(x - \hat{x}_K)^2} = \int_{\hat{x}_K - \frac{\Delta}{2}}^{\hat{x}_K + \frac{\Delta}{2}} (x - \hat{x}_K)^2 P(x) dx \quad (1-3)$$

如果量化值的数目 N 很大, 即量化阶距 Δ 很小, 则抽样值在一个小区间内的概率密度 $P(x)$ 可以看作是常量 $P(\hat{x}_K)$, 即量化值 $\hat{x}_K$ 处的概率密度。将 $P(\hat{x}_K)$ 替代 $P(x)$ 后, 量化噪声功率 N_P 便为

$$N_P = P(\hat{x}_K) \cdot (x - \hat{x}_K)^2 \quad \begin{matrix} \hat{x}_K = \frac{x}{2} \\ \hat{x}_K = \frac{x}{2} \end{matrix}$$

$$= \frac{1}{12} P(\hat{x}_K) \Delta^2 \quad (1-4)$$

在不过载条件下, 即 $|x| \leq x_{max}$ (限幅电平), 量化噪声总功率应为

$$N_{q\Sigma} = \sum_1^N \frac{\Delta^2}{12} P(\hat{x}_K) \Delta = \frac{\Delta^2}{12} \quad (1-5)$$

式中 $N = 2^n$ —— 量化值的数目;

n —— 编码比特数。

由于量化时存在限幅电平 (x_{max}), 对于具有随机性的话音信号, 总有以一定的概率超过限幅电平的信号, 这就形成过载噪声功率, 其平均值为

$$N_o = 2 \int_{x_{max}}^{\infty} (x - x_{max})^2 P(x) dx$$

$$= \sigma_e^2 e^{-\frac{2x_{max}}{\sigma_e}} \quad (1-6)$$

因此, 在有过载情况下, 量化噪声功率为

$$N_q = N_{q\Sigma} + N_o = \frac{\Delta^2}{12} + \sigma_e^2 e^{-\frac{2x_{max}}{\sigma_e}} \quad (1-7)$$

因为均匀量化的量化阶距 Δ 通常选为

$$\Delta = \frac{2x_{max}}{N} \quad (1-8)$$

故式(1-7)可改写成下式:

$$N_q = \frac{x_{max}^2}{3N^2} + \sigma_e^2 e^{-\frac{2x_{max}}{\sigma_e}} \quad (1-9)$$

于是信号量化噪声比为

$$\left(\frac{S}{N_q}\right)_{dB} = 10 \lg\left(\frac{\sigma_s^2}{N_q}\right) = -10 \lg\left[\frac{C^2}{3N^2} + e^{-\sqrt{2}c}\right] \quad (1-10)$$

式中 $c = \frac{x_{max}}{\sigma_e}$ 。若设 $c \gg 10$ ，即在小信号的情况下 ($\sigma_e \ll x_{max}$)， $e^{-\sqrt{2}c}$ 可以略去不计，则信号量化噪声比可近似为

$$\left(\frac{S}{N_q}\right)_{dB} \approx 4.7 + 6n - 20 \lg\left[\frac{x_{max}}{\sigma_e}\right] \quad (1-11)$$

由此可见，每增加一位二进制码 n ，信噪比可以提高 6 dB。若信号有效值 σ_e 减小一倍，即 c 增大一倍，则信噪比便下降 6 dB。

在数字电话通信中，话音信号振幅变化范围很宽，不同讲话人之间以及不同传输条件下，话音信号的功率动态范围约为 40 dB。在这么宽的动态范围内，用均匀量化方法要达到较满意的话音质量，需要用 10~12 位码来编码才能满足要求。

2. 压缩与扩张

均匀量化的优点是简单，但其信号量化噪声比随信号电平的减小而明显下降（在小信号时信噪比很差）。通常希望有一个信噪比与信号电平无关的量化系统，也就是要使误差保持恒定。为了达到这一目的，量化电平必须以对数形式分布，即希望量化器是对输入信号的对数进行量化，而不是对信号本身进行量化。

因为话音信号的概率密度可以近似地用拉普拉斯函数来表示

$$P(x) = \frac{1}{\sigma_e \sqrt{2}} e^{-\frac{\sqrt{2}|x|}{\sigma_e}}$$

从式中可见，出现小幅度的概率是大的。根据典型实验结果，大