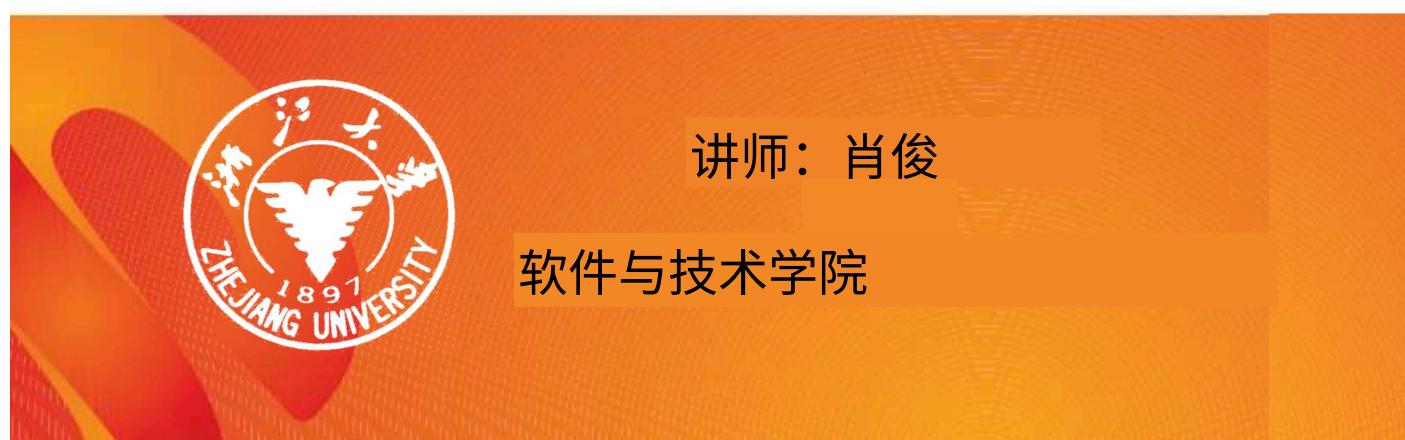


- 声音的数字化
- MIDI：乐器数字接口
- 音频的量化与传输

数字音频基础



1、声音的数字化

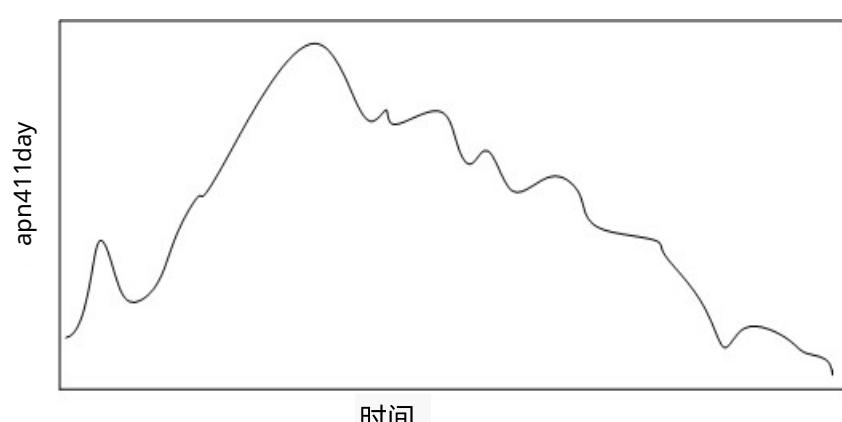
- 什么是声音？
- 数字化
- 奈奎斯特定理
- 信噪比 (SNR)
- 信噪量化比 (SQNR)
- 线性和非线性量化
- 音频滤波
- 音频质量与数据速率
- 合成音效

1.1 什么是声音？

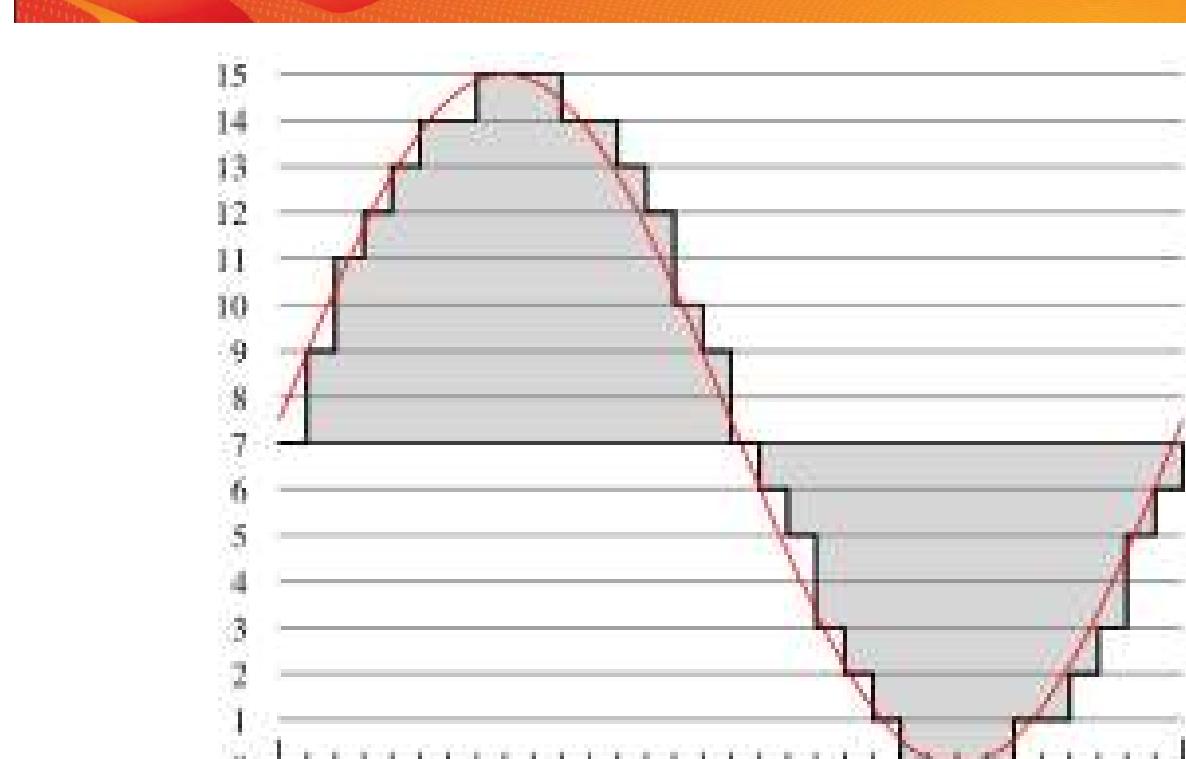
- 声音和光一样，是一种波动现象
- 没有空气——就没有声音
- 声音是一种压力波，具有连续值——声音具有普通波的特性和行为——反射——折射——衍射——可以通过将压力转换为电压水平来测量声音

1.2 数字化

- 数字化是指将其转换为数字流，并且为了提高效率，这些数字最好是整数。



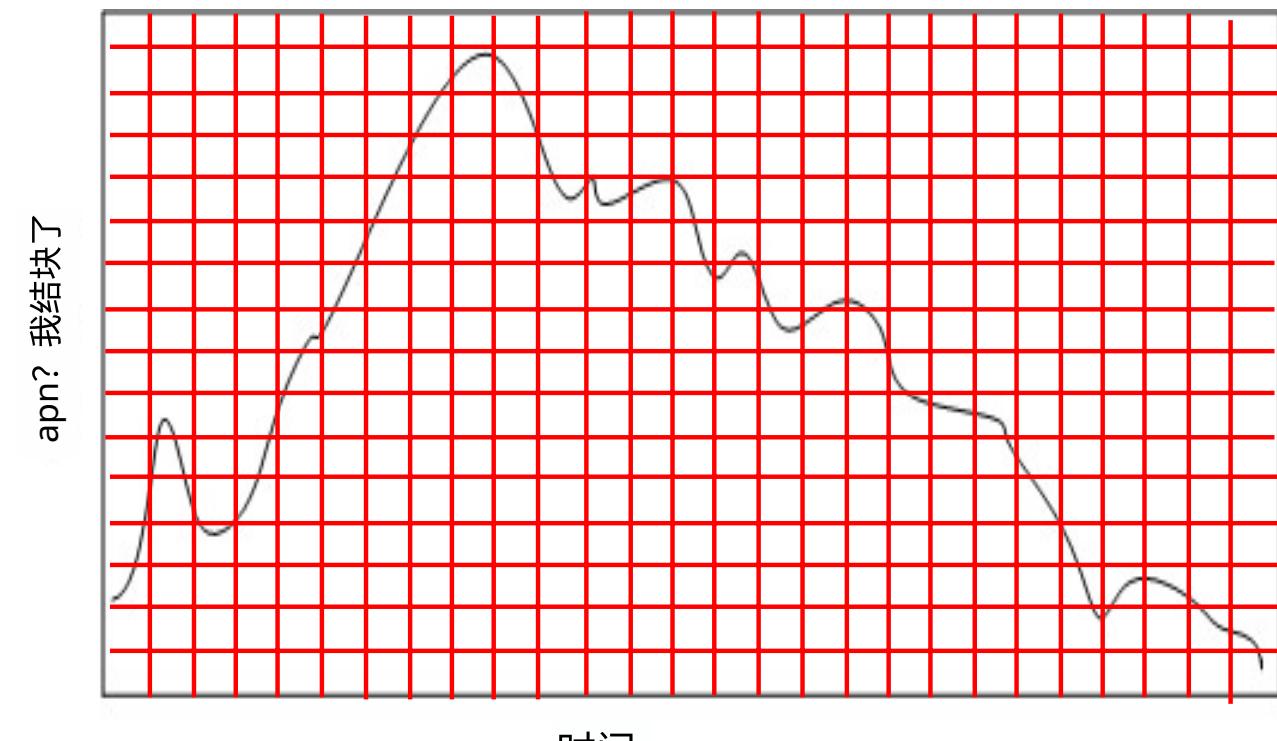
模拟信号：压力波的连续测量



1.2 数字化

- 幅度——连续值且随时间变化——在时间和幅度维度上进行采样
- 时间维度：以均匀间隔进行采样
 - 典型范围：8kHz 到 48kHz，
 - 人类能听到 20 Hz 到 20kHz 的声音
- 量化：在幅度维度上进行采样
 - 均匀采样：等间距采样；
 - 非均匀采样，如 μ 定律规则
 - 典型的均匀量化率：
 - 8位，256个量化级
 - 16位，65536个量化级

1.2 数字化



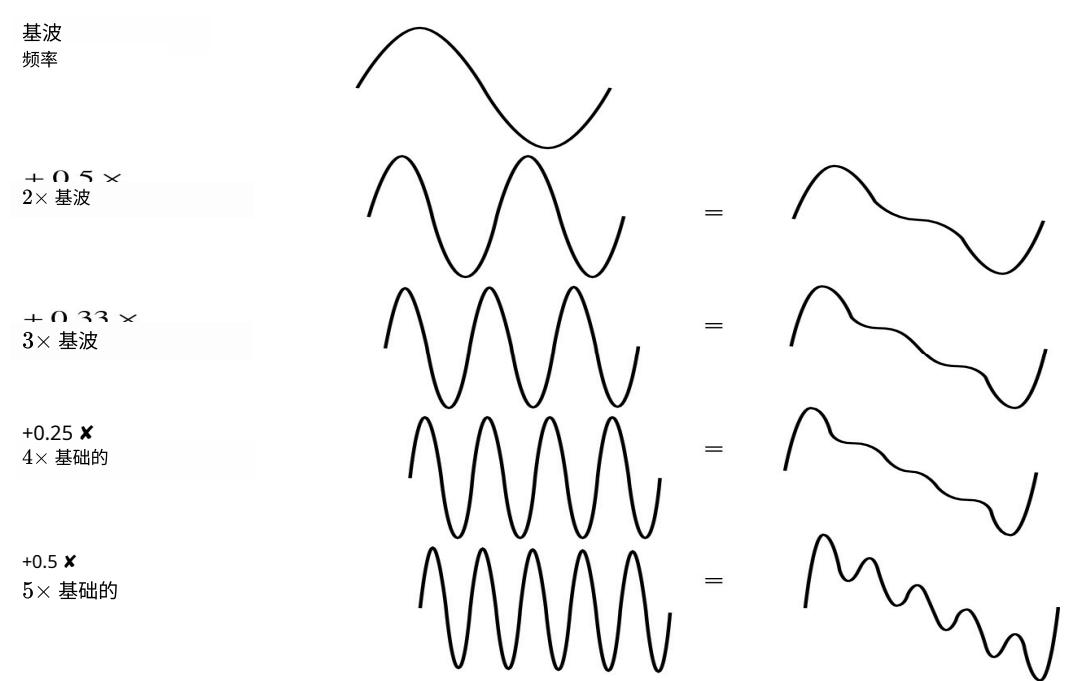
1.2 数字化

- 因此，要决定如何对音频数据进行数字化，我们需要回答以下问题：

- 采样率是多少？
- 数据的量化精度如何，量化是否均匀？
- 音频数据的格式是怎样的？（文件格式）

1.3 奈奎斯特定理

- 信号可以分解为正弦波之和



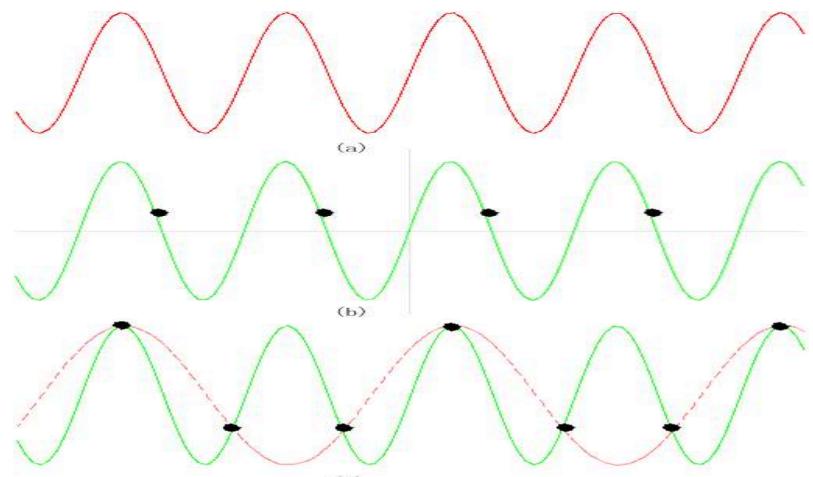
1.3 奈奎斯特定理

•奈奎斯特速率

为了正确采样，采样率必须至少是信号中最大频率成分的两倍——奈奎斯特定理——对于一个带宽受限的信号，其频率下限为 f_1 ，上限为 f_2 ，我们需要至少 $2(f_2 - f_1)$ 的采样率

•奈奎斯特频率：奈奎斯特速率的一半

由于无论如何都不可能恢复高于奈奎斯特频率的频率，因此大多数系统都有一个抗混叠滤波器，它将输入到采样器的频率成分限制在奈奎斯特频率或以下的范围内。



- 单个正弦波
- 通过以实际频率采样检测到的常量
- 通过以 1.5 倍频率采样获得的假信号

1.4 信噪比 (SNR)

正确信号的功率与噪声功率之比称为信噪比 (SNR)
——它是衡量信号质量的一个指标。

信噪比通常以分贝 (dB) 为单位进行测量，其中 dB 是贝尔的十分之一。以 dB 为单位的信噪比数值，是根据电压平方的以 10 为底的对数来定义的，如下所示：

$$SNR = 10 \log_{10} \frac{V_{signal}^2}{V_{noise}^2} = 20 \log_{10} \frac{V_{signal}}{V_{noise}}$$

1.4 信噪比 (SNR)

信号中的功率与电压的平方成正比。例如，如果信号电压 V_{signal} 是噪声的 10 倍，那么信噪比为 $20 * \log_{10}(10) = 20 \text{ dB}$ 。

就功率而言，如果十把小提琴发出的功率是一把小提琴演奏时的 10 倍，那么功率比为 10 dB ，即 1 B 。

须知：功率 — 10；信号电压 — 20。

1.4 信噪比 (SNR)

听觉阈值	0
树叶沙沙声	10
非常安静的房间	20
普通的房间	40
交谈声	60
繁忙的街道	70
响亮的收音机声	80
火车驶过车站的声音	90
铆工	100
不适阈值	120
疼痛阈值	140
鼓膜损伤	160

我们周围常见的声音强度用分贝来描述，它是与我们能听到的最微弱声音的比值。

1.5 SQNR

除了原始模拟信号中可能存在的任何噪声外，量化过程还会产生额外的误差。

(a) 如果电压实际范围在 0 到 1 之间，但我们只有 8 位来存储数值，那么实际上我们会将所有连续的电压值强制转换为仅 256 个不同的值。

(b) 这会引入舍入误差。它并非真正的“噪声”，不过仍被称为量化噪声（或量化误差）。

1.5 SQNR

量化质量由信号量化噪声比 (SQNR) 来表征。

- (a) 量化噪声：在特定采样时刻，模拟信号的实际值与最近量化区间值之间的差值。
- (b) 该误差最大可达区间的一半。
- (c) 对于每个样本 N 位的量化精度，SQNR 可简单表示为：

$$SQNR = 20 \log_{10} \frac{V_{signal}}{V_{quan_noise}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}} = 20 \times N \times \log 2 = 6.02 N (\text{dB})$$

17

1.5 SQNR

(a) 我们将最大信号映射到 $2^{N-1} - 1$ ($\simeq 2^{N-1}$)，将最负信号映射到 -2^{N-1} 。

(b) 式(6.3)是峰值信噪比，即PSQNR：峰值信号和峰值噪声。

(c) 动态范围是信号的最大绝对值与最小绝对值之比： V_{max}/V_{min} 。最大绝对值 V_{max} 映射到 $2^{N-1} - 1$ ；最小绝对值 V_{min} 映射到 1 。 V_{min} 是未被噪声掩盖的最小正电压。最负的信号 $-V_{max}$ 映射到 -2^{N-1} 。

(d) 量化间隔为 $\Delta V = (2 V_{max})/2^N$ ，因为有 2^N 个间隔。整个范围从 V_{max} 到 $(V_{max} - \Delta V/2)$ 被映射到 $2^{N-1} - 1$ 。

(e) 就实际电压而言，最大噪声是量化间隔的一半：
 $\Delta V/2 = V_{max}/2^N$ 。

Fundamentals of Multimedia — Basics of Digital Audio (2024 Spring)

18

1.6 线性和非线性量化

- 线性格式：样本通常存储为均匀量化值
- 考虑到可用比特数有限和人类感知特性
 - 非均匀量化级别更关注人类听觉最佳的频率范围
 - 非均匀量化方案利用了人类的感知特性并使用对数。

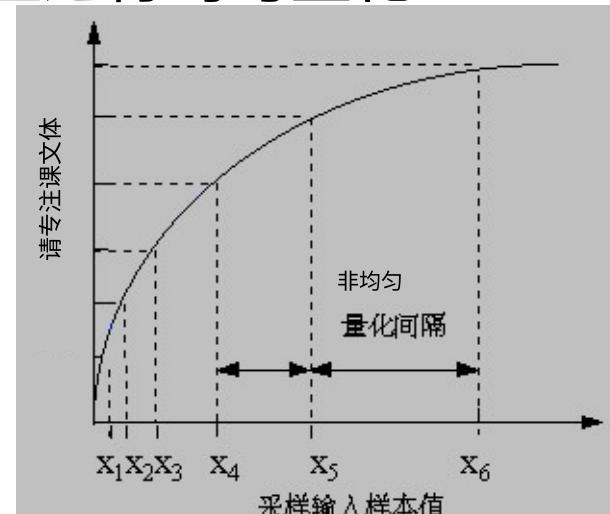
19

1.6 线性和非线性量化

• 非线性量化步骤

- 将模拟信号从原始S空间转换到理论 R 空间；

• 对所得值进行均匀量化



多媒体基础

数字音频基础 (2024年春季)

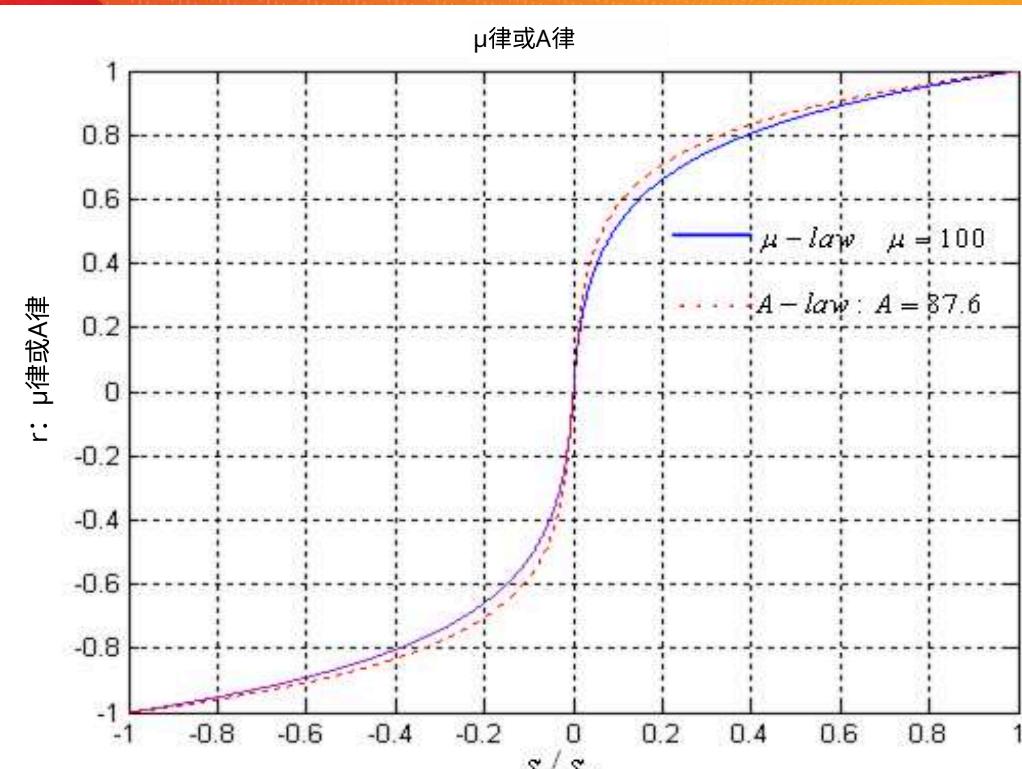
20

1.6 线性和非线性量化

- 方程如下：
 $\mu\text{律 } r = \frac{\text{sgn}(s)}{\ln(1+\mu)} \ln \left\{ 1 + \mu \left| \frac{s}{s_p} \right| \right\}, \left| \frac{s}{s_p} \right| \leq 1$
 $A\text{律 } r = \begin{cases} \frac{A}{1+\ln A} \left(\frac{s}{s_p} \right), & \text{其中 } \text{sgn}(s) \begin{cases} 1 & \text{if } s > 0 \\ -1 & \text{otherwise} \end{cases} \\ \frac{\text{sgn}(s)}{1+\ln A} \left(1 + \ln A \frac{s}{s_p} \right), & \frac{A}{1+\ln A} \leq \frac{s}{s_p} \leq 1 \end{cases}$
 $-\mu = 100$ 或 255
 $-A = 87.6$
• s/s_p 的范围是 -1 到 1

21

1.6 线性和非线性量化



Fundamentals of Multimedia — Basics of Digital Audio (2024 Spring)

22

1.7 音频滤波

- 在采样和模数转换之前，通过对音频信号进行滤波来去除不需要的频率
- 保留的频率取决于应用场景：
 - 语音，包含 50Hz 至 10kHz；
 - 音频音乐信号，包含 20Hz 至 20kHz
- 其他频率被带通滤波器（也称为限带滤波器）阻挡；

23

1.8 音频质量与数据速率

- 随着用于量化的比特数增多，未压缩数据速率会增加
- 音频质量——数据速率和带宽
 - 模拟设备，带宽以频率单位赫兹表示；
 - 数字设备，以每秒比特数 (bps) 表示

Fundamentals of Multimedia — Basics of Digital Audio (2024 Spring)

24

1.8 音频质量与数据速率

质量	采样速率 (千赫兹)	每采样位数	单声道/立体声	数据速率 (未压缩) (千字节/秒)	频段 (千赫兹)
电话	8	8	单声道	8	0.200-3.4
调幅广播	11.025	8	单声道	11.0	0.1-5.5
调频广播	22.05	16	立体声	88.2	0.02-11
CD	44.1	16	立体声	176.4	0.005-20
DAT	48	16	立体声	192.0	0.005-20
DVD音频	192 (最大值)	24 (最大值)	6通道	1200 (最大值)	0 - 96 (最大值)

1.9 合成音效

•两种方法：将数字化声音转换为模拟声音

•调频 - 频率调制；

•波表 (更精确)

•数字样本是真实乐器的存储音效

•在调频中，通过添加涉及第二个调制频率的另一项来改变载波正弦波。

$X(t) = A(t) \cos [\omega_c \pi t + I(t) \cos (\omega_m \pi t + \phi_m) + \phi_c]$
- A(t) 包络，声音的响度； - I(t)，通过改变调制频率产生和声感；

• ϕ_c 和 ϕ_m 相位常数，产生时移

1.9 合成声音

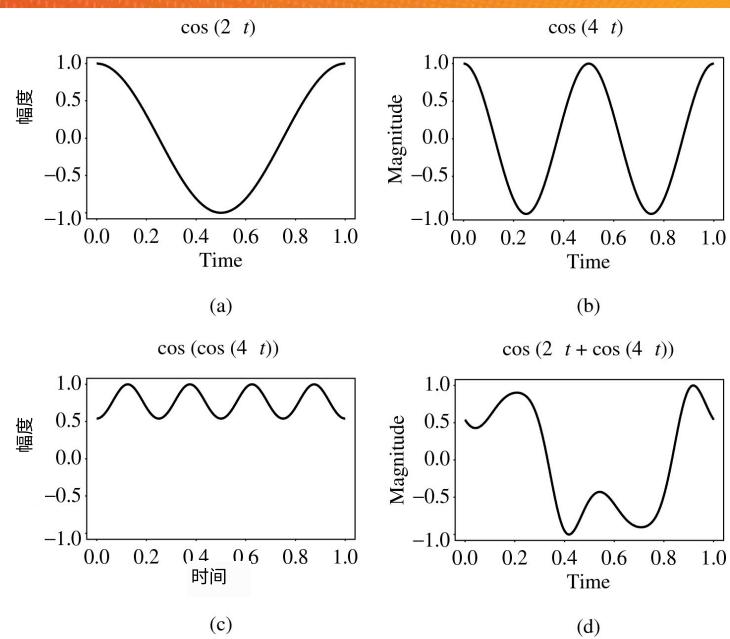
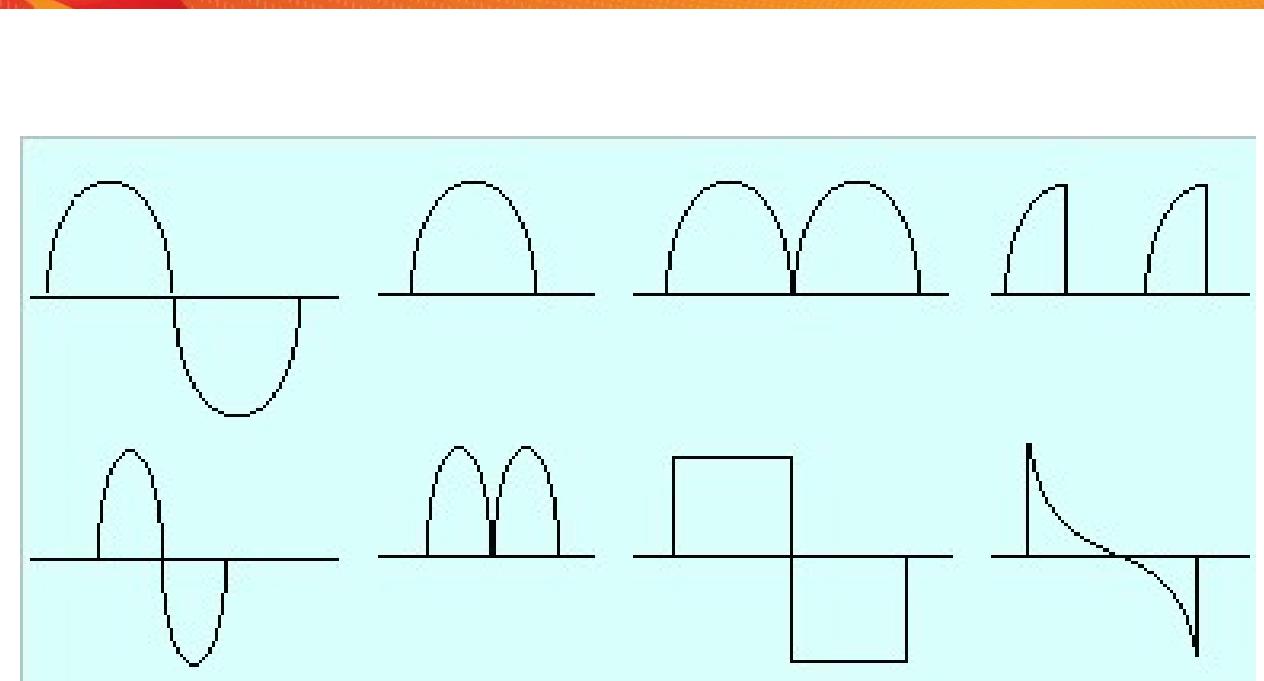


图6.7：频率调制。(a): 单一频率。(b): 两倍频率。(c): 通常，调频是通过将正弦函数的自变量设为另一个正弦函数来实现的。(d): 更复杂的形式由载波频率 $2\pi t$ 和正弦函数内的调制频率 $4\pi t$ 余弦函数产生。



1.9 合成音效

波表合成：一种从数字信号生成声音的更精确方法。
也简称为采样。

在这种技术中，存储了真实乐器声音的实际数字样本。由于波表存储在声卡的内存中，因此可以通过软件对其进行操作，从而实现声音的组合、编辑和增强。

→ 详情链接。

2、乐器数字接口

- MIDI

•术语

•MIDI与MP3的区别

2.1 MIDI

MIDI (乐器数字接口)

□ 一种使音乐设备能够相互通信的协议

•MIDI - 脚本语言

•不是音频信号，而是发送到MIDI设备以产生声音或执行某些操作的指令序列

•MIDI生成音乐的方法 - 调频 (FM)

•波表合成

2.2 术语

合成器：

•声音发生器 - 改变音高、响度、音色

•微处理器、键盘、控制面板、内存等。

音序器：

•用于编辑一系列音乐事件的硬件或软件

•一个或多个MIDI输入 (Ins) 和输出 (OUTs)。

2.2 术语

通道：

- 独立的MIDI消息
- 16个通道
(与16种乐器相关联)

音色：

- 声音品质，例如钢琴、小提琴等。
- 多音色——可同时演奏多种不同声音（例如钢琴、铜管乐器、鼓等）

2.3 MIDI与MP3的区别

• MIDI文件——指令的集合

- 非常小，通常约为10k大小，而MP3文件通常超过2兆字节

• MP3——音质与CD相近

- MIDI只能生成简单的音乐音调，无法还原歌手的声音。
- 许多软件可以将mp3转换为midi格式，如widi。

3、音频的量化与传输

- 音频编码
- 脉冲编码调制
- 音频差分编码
- 无损预测编码
 - DPCM
 - DM
 - ADPCM

3.1 音频编码

- 编码——数据的量化与变换
- 利用音频信号中的时间冗余

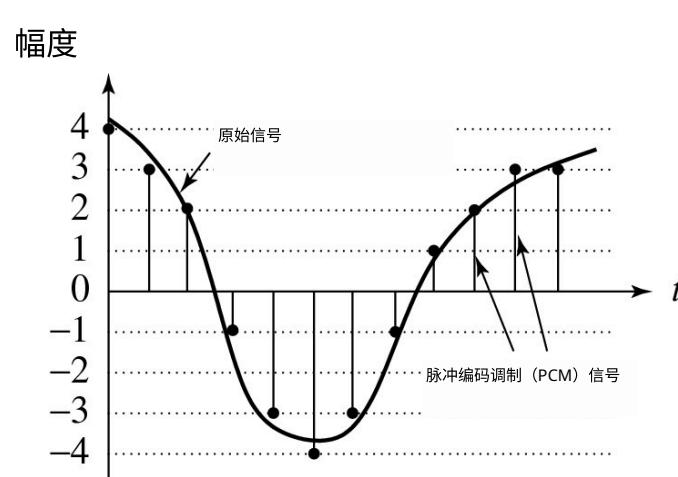
• 减小信号值的大小

• 当前信号与过去某一时刻信号之间的差异可以减小信号值的大小，还能将像素值（当前差异值）的直方图集中到一个小得多的范围内。

• 无损压缩方法可以产生更短的比特长度；-为音频生成量化输出 - PCM，脉冲编码调制 - DPCM，PCM的差分版本 - ADPCM，自适应DPCM

3.2 脉冲编码调制

- 从模拟信号创建数字信号的基本技术是采样和量化。
- 量化包括选择幅度上的断点，然后将区间内的任何值重新映射到一个代表性输出电平上。



3.2 脉冲编码调制

a) 区间边界的集合称为决策边界，代表值称为重建电平。

b) 所有将被映射到同一输出电平的量化器输入区间的边界构成编码器映射。

c) 作为量化器输出值的代表值构成解码器映射。

d) 最后，我们可能希望通过为最常见的信号值分配使用更少比特的比特流来压缩数据（第7章）。

3.2 脉冲编码调制

每个压缩方案都有三个阶段：

- 将输入数据转换为一种新的表示形式，以便更轻松或更高效地进行压缩。
- 我们可能会引入信息损失。量化是主要的有损步骤
⇒ 我们使用有限数量的重建级别，比原始信号中的级别更少。
- 编码。为每个输出级别或符号分配一个码字（从而形成一个二进制比特流）。这可以是定长码，也可以是诸如霍夫曼编码（第7章）之类的变长码。

3.2 脉冲编码调制

• 语音压缩中的PCM

• 假设语音带宽约为50 Hz至约10kHz，奈奎斯特速率要求采样率为20kHz。

(a) 使用无压扩的均匀量化，我们能采用的最小样本大小可能约为12位。因此，对于单声道语音传输，比特率将为240kbps。

(b) 通过压扩，我们可以在相同感知质量水平下将样本大小降至约8位，从而将比特率降至160kbps。

(c) 然而，电话通信的标准方法实际上假设我们要重现的最高频率音频信号仅约为4kHz。因此，采样率仅为8kHz，经过压扩后的比特率因此降至64kbps。

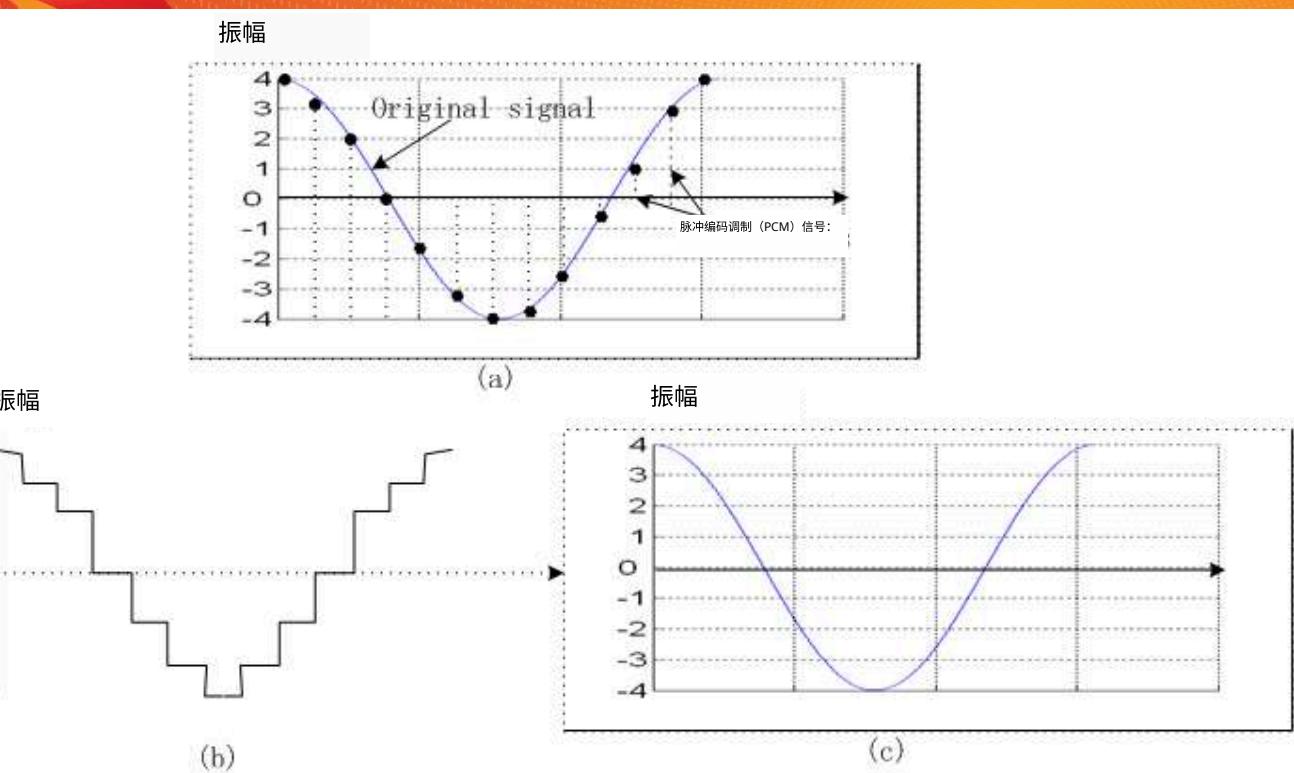
3.2 脉冲编码调制

然而，我们还必须解决两个小问题：

- 由于只考虑频率高达 4kHz 的声音，所有其他频率成分必定是噪声。因此，我们应该从模拟输入信号中去除这种高频成分。这可以通过使用一个带限滤波器来实现，该滤波器可以阻挡高频以及极低频率。

此外，一旦我们得到一个脉冲信号，如下方图 6.13(a) 所示，我们仍然必须进行数模转换，然后构建最终的输出模拟信号。但实际上，我们得到的信号是图 6.13(b) 所示的阶梯信号。

3.2 脉冲编码调制



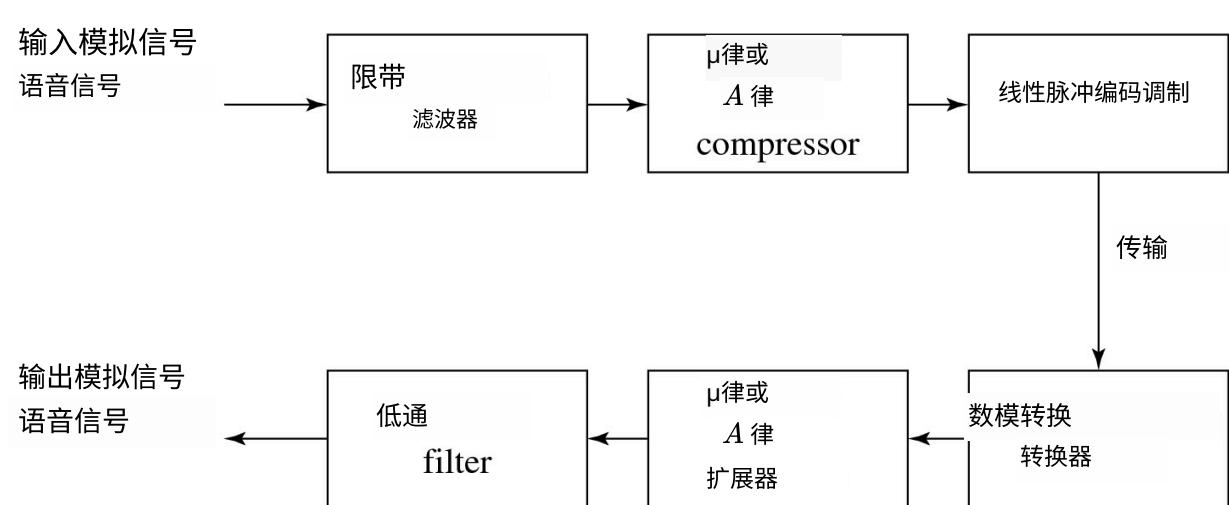
3.2 脉冲编码调制

- 一个不连续信号不仅包含原始信号的频率分量，还包含理论上无限的高频分量集：

- 这一结果源于信号处理中的傅里叶分析理论。
- 这些高频分量是多余的。
- 因此，数模转换器的输出会进入一个低通滤波器，该滤波器只允许保留原始信号的最高频率。

3.2 脉冲编码调制

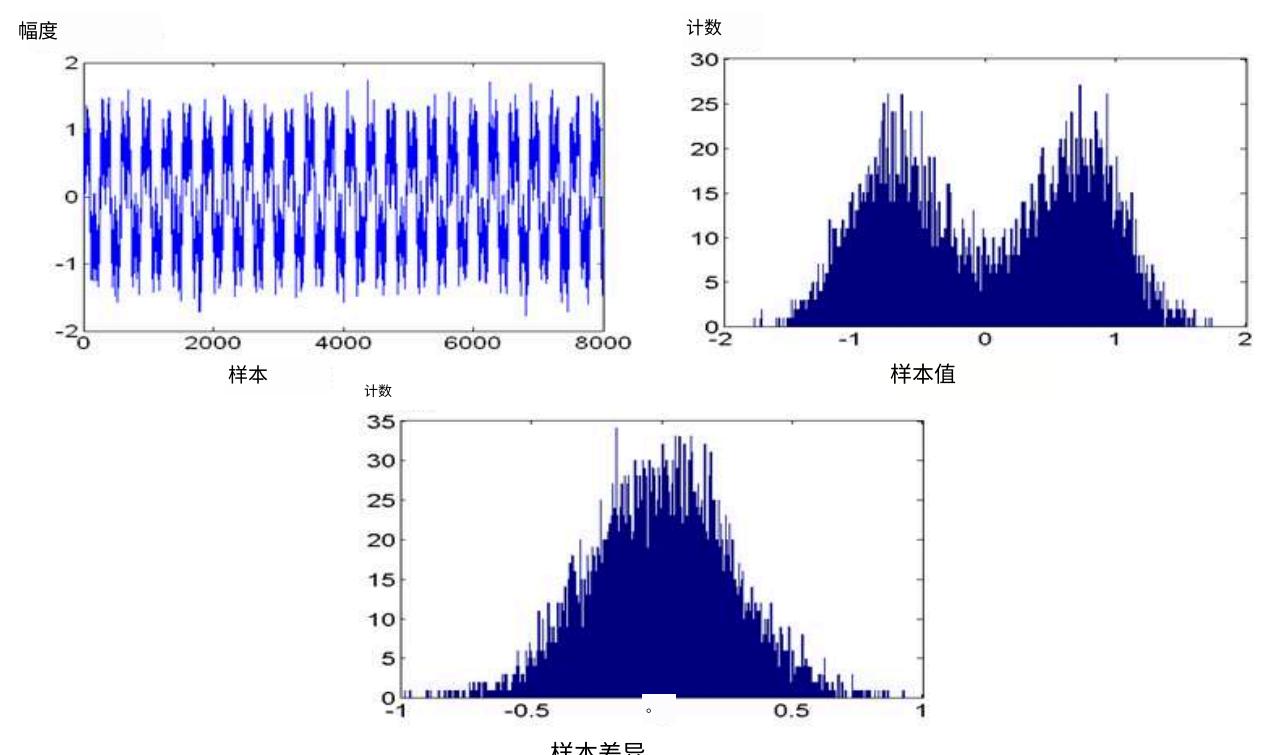
- 电话信号的完整编码和解码方案如图6.14所示。经过低通滤波后，输出变得平滑，上述图6.13(c)展示了这种效果。



3.3 音频的差分编码

- 音频通常不是以简单的脉冲编码调制 (PCM) 形式存储，而是以一种利用差值的形式存储 — 差值通常是较小的数字，因此有可能使用更少的比特来存储。
- 如果一个随时间变化的信号在时间上具有一定的连贯性（“时间冗余”），那么通过用当前样本减去前一个样本得到的差值信号，其直方图会更集中，最大值会在零附近。

3.3 音频的差分编码



3.4 无损预测编码

- 预测编码：简单来说就是传输差值——将下一个样本预测为与当前样本相等；不传输样本本身，而是传输前一个样本和下一个样本之间的差值。
 - 预测编码包括找出差值，并使用脉冲编码调制 (PCM) 系统传输这些差值。
 - 注意，整数的差值仍为整数。将整数输入信号表示为值的集合 f_n 。然后我们简单地将值 f 预测为前一个值，并将误差 e_n 定义为实际信号与预测信号之间的差值：

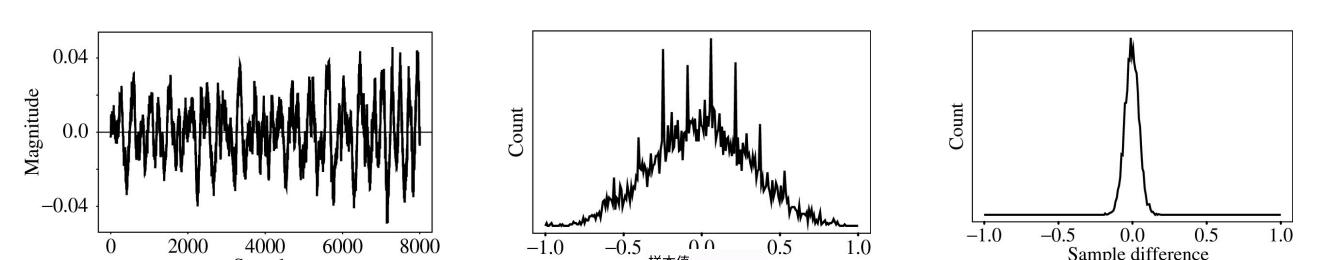
$$\hat{f}_n = f_{n-1} \quad e_n = f_n - \hat{f}_n \quad (6.12)$$

3.4 无损预测编码

- 但通常情况下，前几个值（如 f_{n-1} 、 f_{n-2} 、 f_{n-3} 等）的某种函数能提供更好的预测。通常会使用线性预测函数：

$$\hat{f}_n = \sum_{k=1}^{2 to 4} a_{n-k} f_{n-k}$$

求差分的目的是使样本值的直方图更尖峰化。



3.4 无损预测编码

一个问题：假设我们的整数样本值范围在0到255之间。那么差值可能在 -255 到 255 之间——我们将动态范围（最大值与最小值之比）扩大了两倍 →，传输某些差值需要更多的比特位。

(a) 针对此问题的一个巧妙解决方案：定义两个新代码，分别表示为 SU 和 SD，代表上移和下移。将为这些代码保留一些特殊的代码值。

(b) 然后，我们可以仅针对有限的信号差值集合使用码字，例如仅使用 -15 到 16 的范围。处于该有限范围内的差值可以按原样编码，但需使用

对于 SU、SD 的额外两个值，范围 -15..16 之外的值可以作为一系列偏移量进行传输，随后是一个确实在 -15..16 范围内的值。

(c) 例如，100 以下形式传输：SU, SU, SU, 4，其中 SU 和 4 的（编码）是被传输（或存储）的内容。

3.4 无损预测编码

•无损预测编码——解码器产生的信号与原始信号相同。举个简单的例子，假设我们为 \hat{f}_n 设计了如下预测器：

$$\hat{f}_n = \left\lfloor \frac{1}{2}(f_{n-1} + f_{n-2}) \right\rfloor$$

$$e_n = f_n - \hat{f}_n$$

3.4 无损预测编码

•让我们考虑一个具体的例子。假设我们要对序列 $f_1, f_2, f_3, f_4, f_5 = 21, 22, 27, 25, 22$ 进行编码。为了使用预测器，我们将创建一个额外的信号值 f_0 ，其等于 $f_1 = 21$ ，并首先无编码地传输这个初始值：

$$\hat{f}_2 = 21, e_2 = 22 - 21 = 1;$$

$$\hat{f}_3 = \left\lfloor \frac{1}{2}(f_1 + f_2) \right\rfloor = \left\lfloor \frac{1}{2}(22 + 21) \right\rfloor = 21, \\ e_3 = 27 - 21 = 6;$$

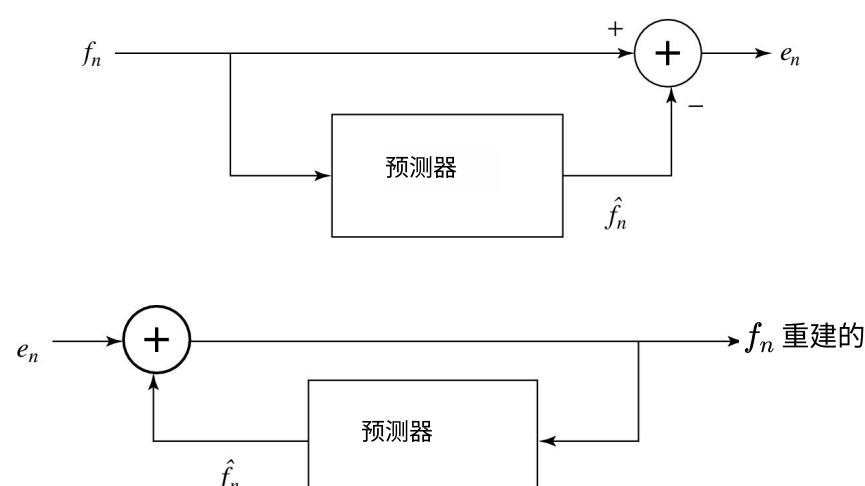
$$\hat{f}_4 = \left\lfloor \frac{1}{2}(f_2 + f_3) \right\rfloor = \left\lfloor \frac{1}{2}(27 + 22) \right\rfloor = 24, \\ e_4 = 25 - 24 = 1;$$

$$\hat{f}_5 = \left\lfloor \frac{1}{2}(f_3 + f_4) \right\rfloor = \left\lfloor \frac{1}{2}(25 + 27) \right\rfloor = 26,$$

$$e_5 = 22 - 26 = -4$$

3.4 无损预测编码

•我们可以看到，误差确实以零为中心，并且编码（分配位串码字）将是高效的。图6.16展示了用于封装此类系统的典型示意图：



3.5 DPCM

•差分脉冲编码调制（DPCM）与预测编码完全相同，只是它包含一个量化器步骤。

$$\hat{f}_n = function_of(\tilde{f}_{n-1}, \tilde{f}_{n-2}, \tilde{f}_{n-3}, \dots)$$

$$e_n = f_n - \hat{f}_n$$

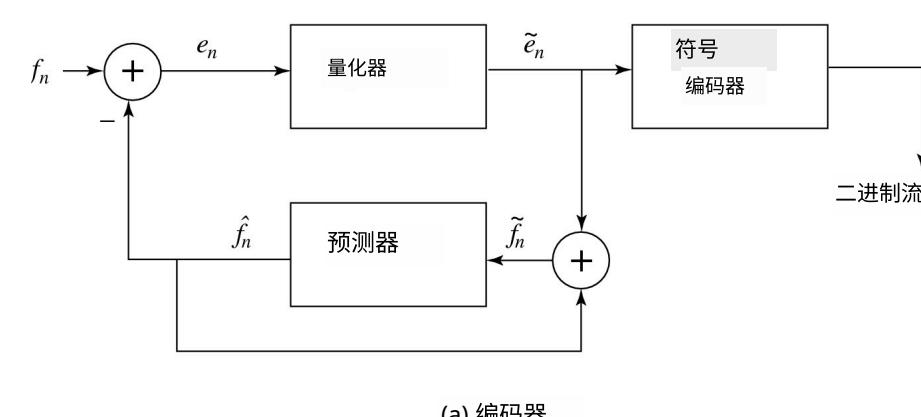
$$\tilde{e}_n = Q[e_n]$$

传输码字 (\tilde{e}_n)

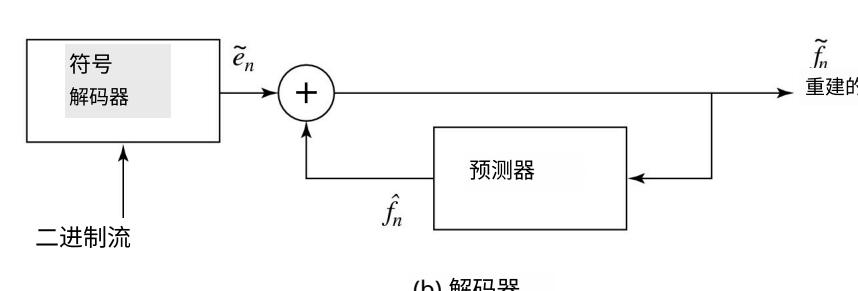
$$reconstruct: \tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

然后是.....的码字
量化误差值
使用.....生成
熵编码，例如
霍夫曼编码
(第7章)

3.5 DPCM



(a) 编码器



(b) 解码器

3.5 DPCM

•注意，量化噪声 $f_n - \hat{f}_n$ 等于误差项上的量化效应 $e_n - \tilde{e}_n$ 。
•让我们来看实际的数字：假设我们采用以下特定的预测器：

$$\hat{f}_n = trunc(\tilde{f}_{n-1} + \tilde{f}_{n-2}) \quad (6.19)$$

这样 $e_n = f_n - \hat{f}_n$ 就是一个整数。

•同样，使用量化方案：

$$\tilde{e}_n = Q[e_n] = 16 * trunc((255 + e_n) / 16) - 256 + 8$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n \quad (6.20)$$

3.5 DPCM

•首先，我们注意到误差范围在 -255 到 255 之间，即误差项有 511 个可能的电平。量化器只是将误差范围划分为 32 个区间，每个区间约有 16 个电平。它还使每个区间的代表性重构值等于每组 16 个电平的中点。

e, 在范围内	量化为值
-255 .. -240	-248
-239 .. -224	-232
.	.
-31 .. -16	-24
-15 .. 0	-8
1 .. 16	8
17 .. 32	24
.	.
225 .. 240	232
241 .. 255	248

3.5 DPCM

- 作为信号值流的一个示例，考虑以下值集合：

$$\begin{array}{ccccc} f_1 & f_2 & f_3 & f_4 & f_5 \\ 130 & 150 & 140 & 200 & 300 \end{array}$$

- 在前面添加额外的值 $f = 130$ 以复制第一个值。用量化误差 $\tilde{e}_1 \equiv 0$ 进行初始化，这样第一个重构值就是精确的： $\tilde{f}_1 = 130$ 。然后计算得到的其余值如下（前面添加的值用方框框起来）：

\hat{f}	130	130, 142, 144,
e	0	16720, -2, 56, 63
\tilde{e}	0	24, -8, 56, 56
\tilde{f}	130	154, 134, 200, 223

- 在解码器端，我们再次假设额外的值 $\sqrt[3]{\tilde{f}}$ 等于正确值 f_1 ，这样第一个重构值 \tilde{f}_1^J 就是正确的。接收到的是 \tilde{e} ，并且如果我们使用完全相同的预测规则，重构的 \tilde{f}_1 与编码器端的相同。

3.6 DM

- 增量调制（DM）：差分脉冲编码调制（DPCM）的简化版本。常被用作快速模数转换器。

- 均匀增量调制：仅使用单个量化误差值，可为正或负。

(a) \Rightarrow 一种1位编码器。以阶梯方式生成跟随原始信号的编码输出。方程组如下：

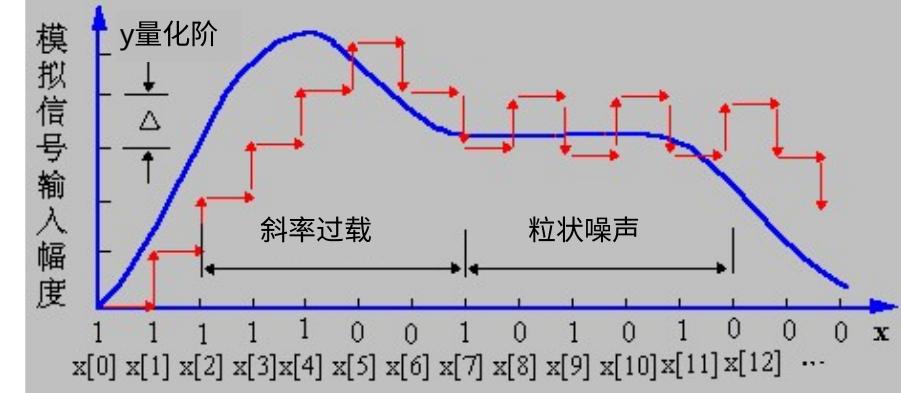
$$\begin{aligned}\hat{f}_n &= \tilde{f}_{n-1}, \\ e_n &= f_n - \hat{f}_n = f_n - \tilde{f}_{n-1},\end{aligned}$$

$$\tilde{e}_n = \begin{cases} +k & \text{if } e_n > 0, \text{ where } k \text{ is a constant} \\ -k & \text{otherwise} \end{cases}$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n.$$

3.6 DM

- 自适应增量调制（DM）：如果实际信号曲线的斜率较大，阶梯近似就无法跟上。对于陡峭的曲线，应自适应地改变步长 k 。



自适应差分调制（Adaptive DM）：简单地自适应改变步长 k ，即根据信号的当前特性进行调整。

3.6 ADPCM

- 自适应差分脉冲编码调制（Adaptive DPCM），使编码器进一步适应输入信号。

- 调整量化器步长以适应输入信号：

- 利用输入信号的特性；前向自适应量化

- 利用量化输出的特性；后向自适应量化

- 自适应预测编码：改变预测系数

- 如果我们使用 M 个先前的值，那么 M 个系数 $a_i, i = 1..M$
- 使用最小二乘法，求出 a_i 的最佳值：

$$\hat{f}_n = \sum_{i=1}^M a_i \tilde{f}_{n-i} \quad \min \sum_{n=1}^N (f_n - \hat{f}_n)^2$$

结束

谢谢！

邮箱：junx@cs.zju.edu.cn