



数字信号处理基础知识

杨榆，钮心忻，周琳娜，雷敏

北京邮电大学信息安全中心

yangyu@bupt.edu.cn

音频信号处理基础

- 数字音频的概念
- 人类听觉系统(HAS)特点
- 语音质量评价
- 语音信号处理基础

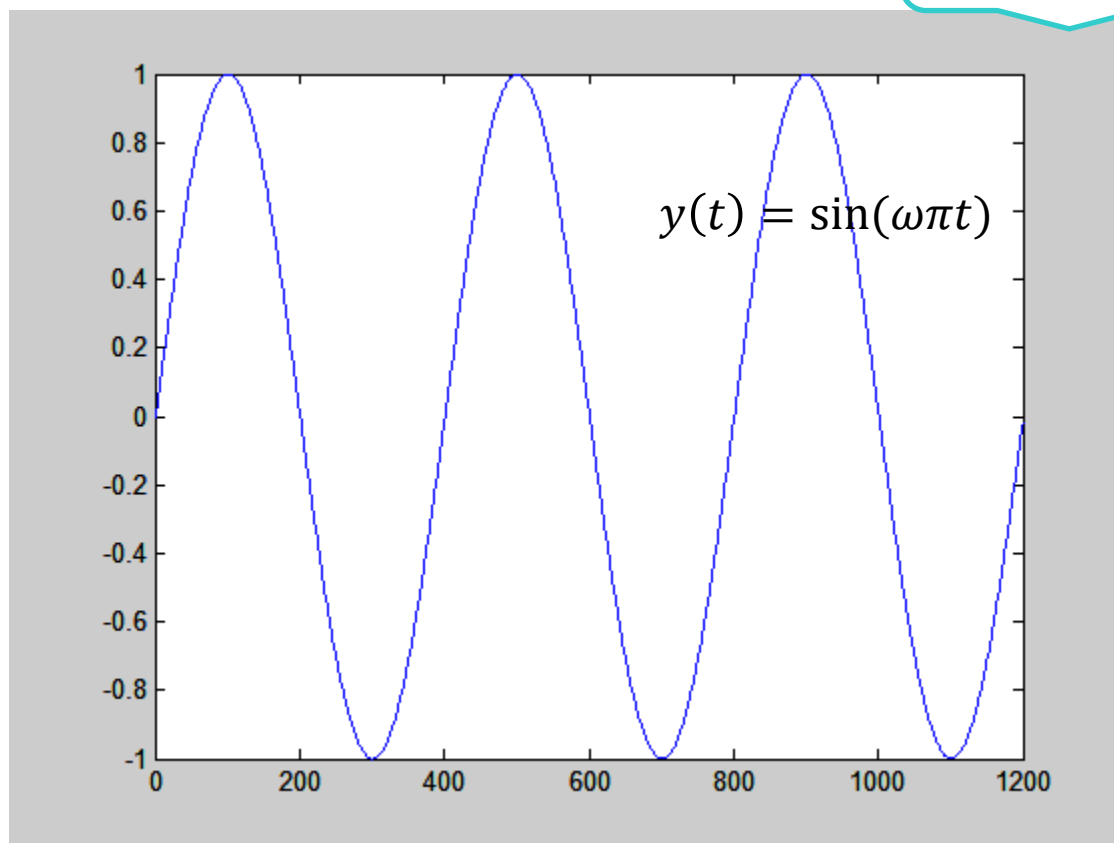
音频基础——数字音频概念

○ 什么是模拟信号？

- 现实生活中的语音信号，从时间及其大小来看，都是连续的。

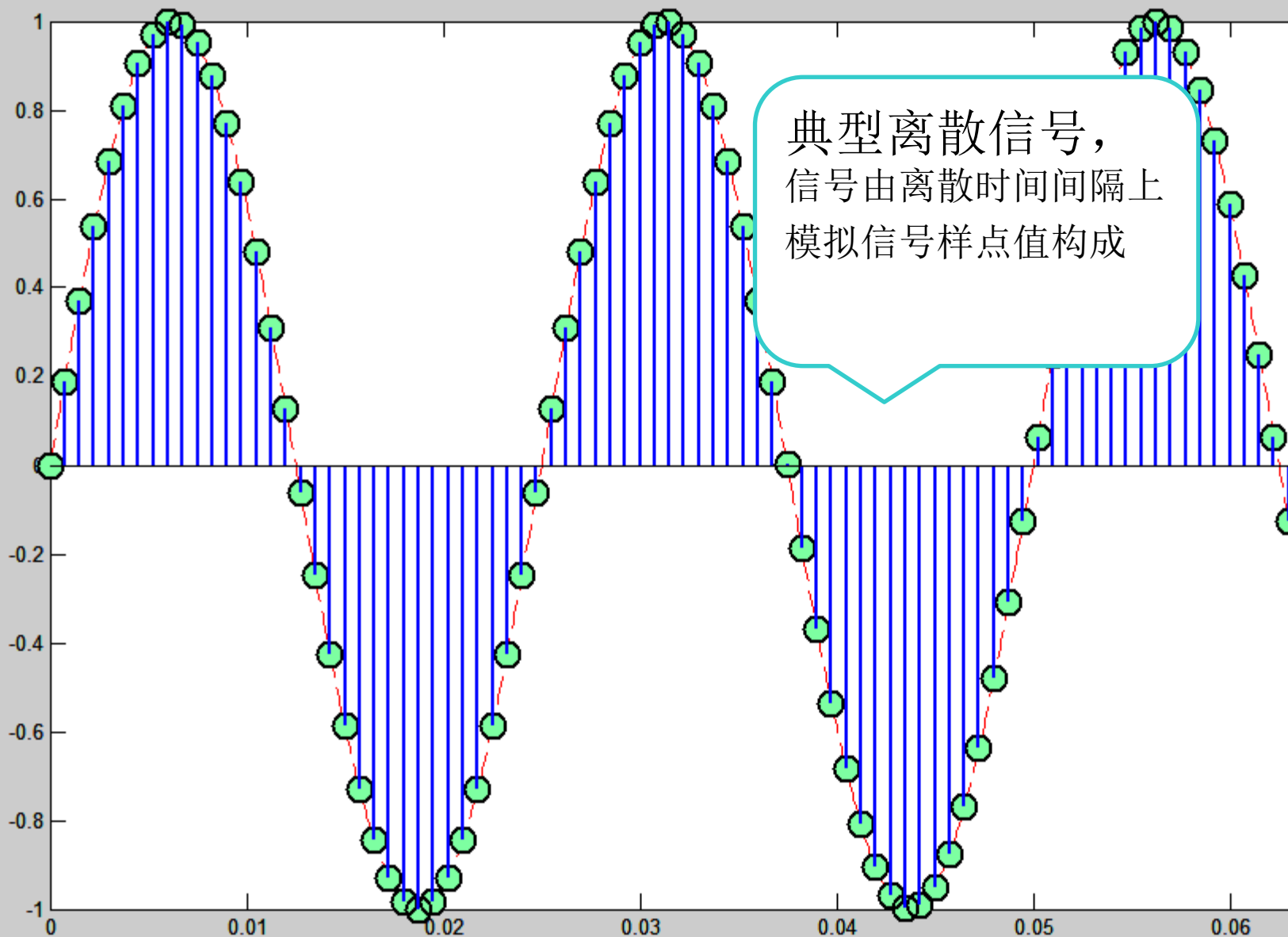
音频基础——数字音频概念

典型模拟信号形态

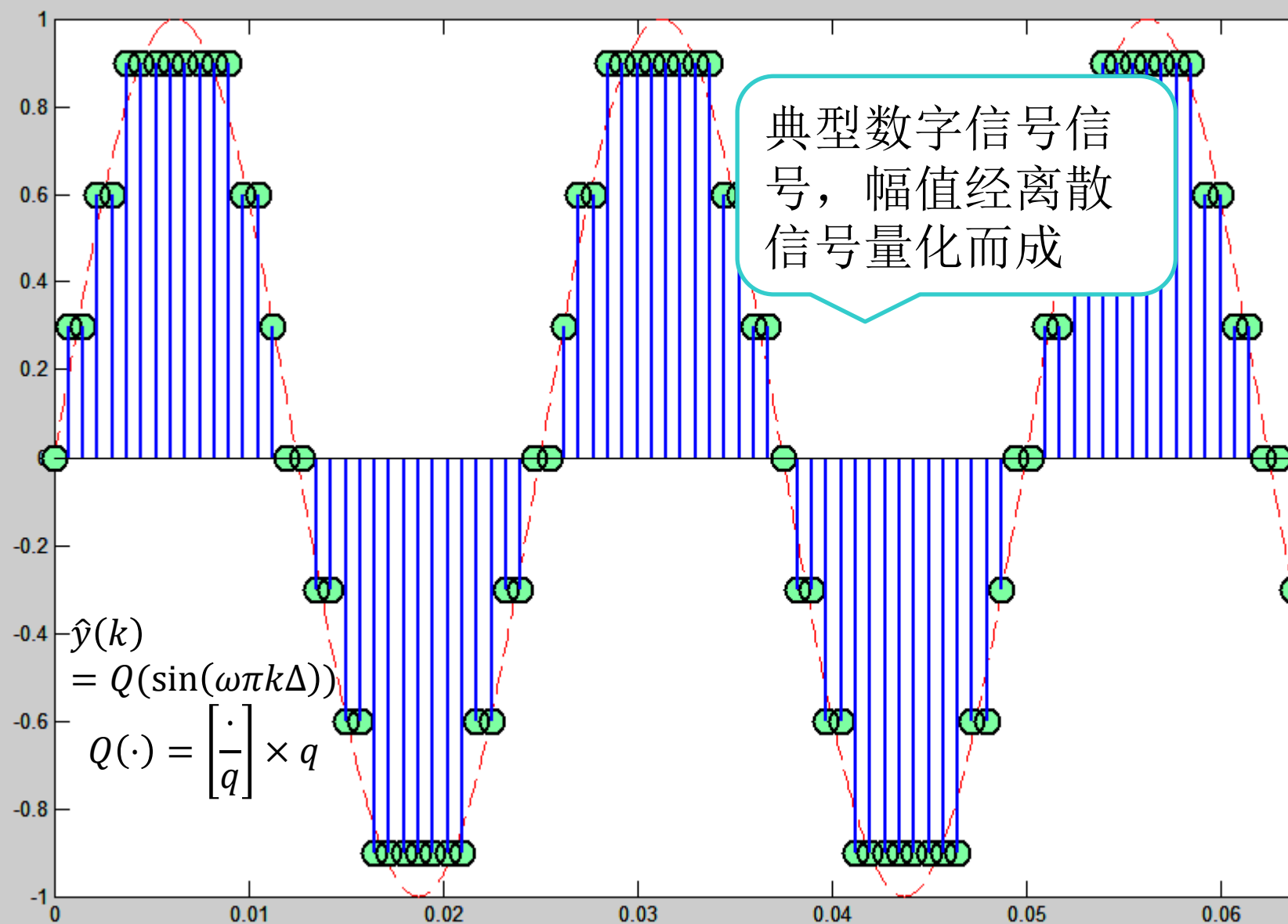


音频基础——数字音频概念

- 什么是数字信号？
 - 数字信号是计算机能表示，能存储的信号。
 - 其时间和大小的精度都是有限的。
- 如何用数字信号表示模拟信号？
 - 每隔一定时间间隔观察一次信号大小，并量化表示。



$$y(k) = \sin(\omega\pi k\Delta) = 0.00, 0.19, 0.37, 0.54, 0.68, 0.81, 0.90, 0.97...$$



音频基础——数字音频概念

○ 采样频率

- “观察”信号的频度，单位为Hz。

○ 码速率

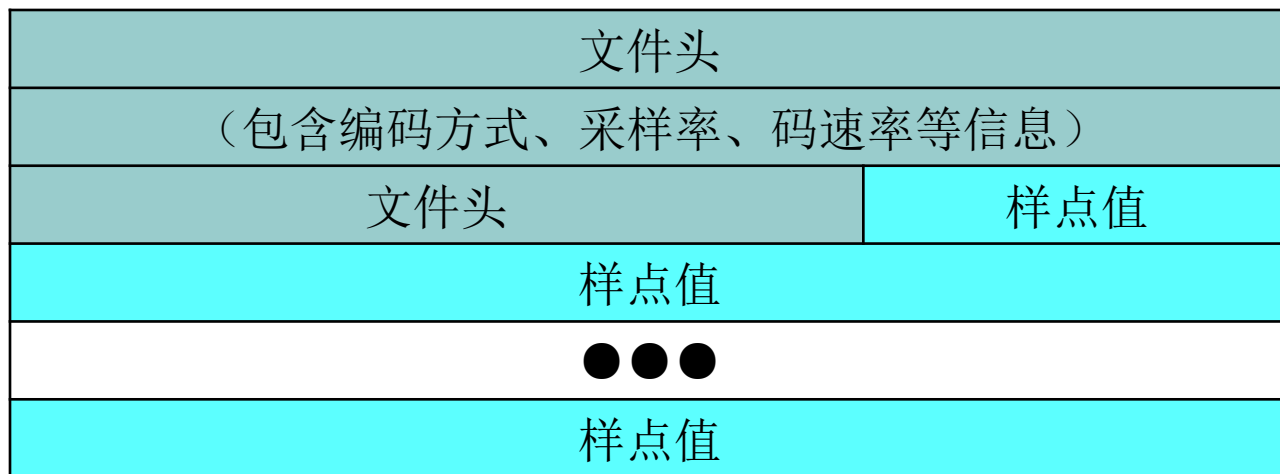
- 我们需要对采样所得信息进行编码，产生的码流单位为b/s(bit per second)。
- PCM码速率：64K、128K、256K b/s

音频基础——数字音频概念

- 为何数字信号能表示模拟信号？
 - 奈奎斯特采样定律告诉我们，当采样频率满足一定要求时，数字信号能完整保留原始信号信息。
 - 典型采样频率
 - 话音：8kHz；音乐：22.05kHz，或44.1kHz
 - 并且，采样值的编码方式是依据声音心理学模型设计，编解码所引起的失真是可接受的。

音频基础——数字音频文件

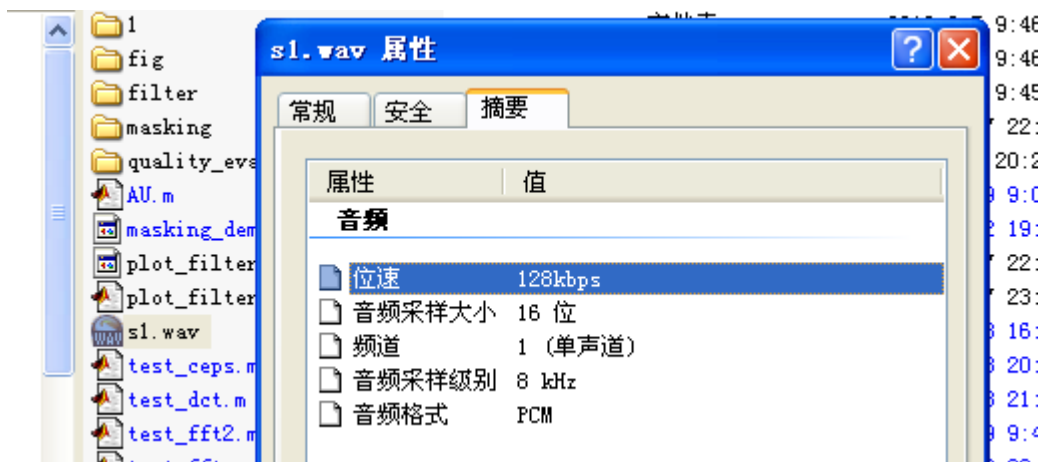
- 音频文件格式
 - WAV文件格式



音频基础——数字音频文件

○ 音频文件实例

- WAV文件: s1.wav (Fs:8KHz, Nbits=16)



音频基础——数字音频文件

○ 音频文件实例

- WAV文件: s1.wav (Fs:8KHz, Nbits=16)

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f	
00000000h:	52	49	46	46	24	E8	07	00	57	41	56	45	66	6D	74	20	; RIFF\$?.WAVEfmt
00000010h:	10	00	00	00	01	00	01	00	40	1F	00	00	80	3E	00	00	;@...€>.
00000020h:	02	00	10	00	64	61	74	61	00	E8	07	00	F4	FE	B2	FE	;data.?.酷猖
00000030h:	1B	FE	94	FD	CA	FD	2B	FE	BC	FE	82	FF	50	FF	66	FF	; .簪 ? 贻 P...

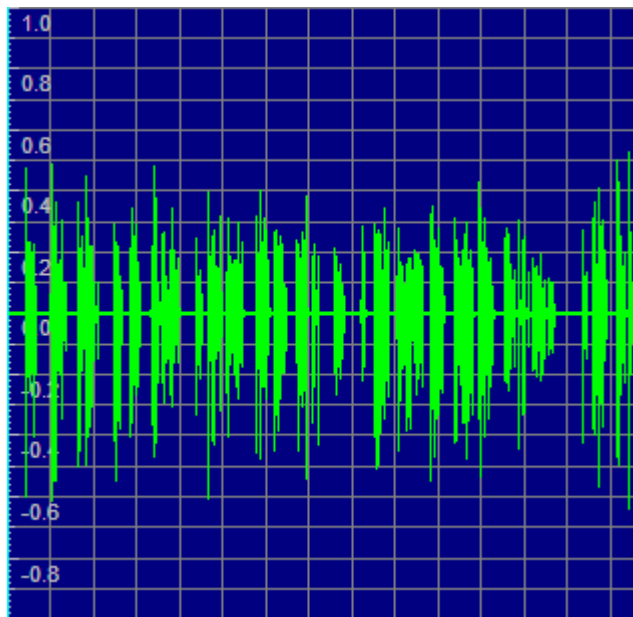
前44个字节为音频文件头

第45、46个字节为第一个
样点量化幅值

音频基础——单双声道

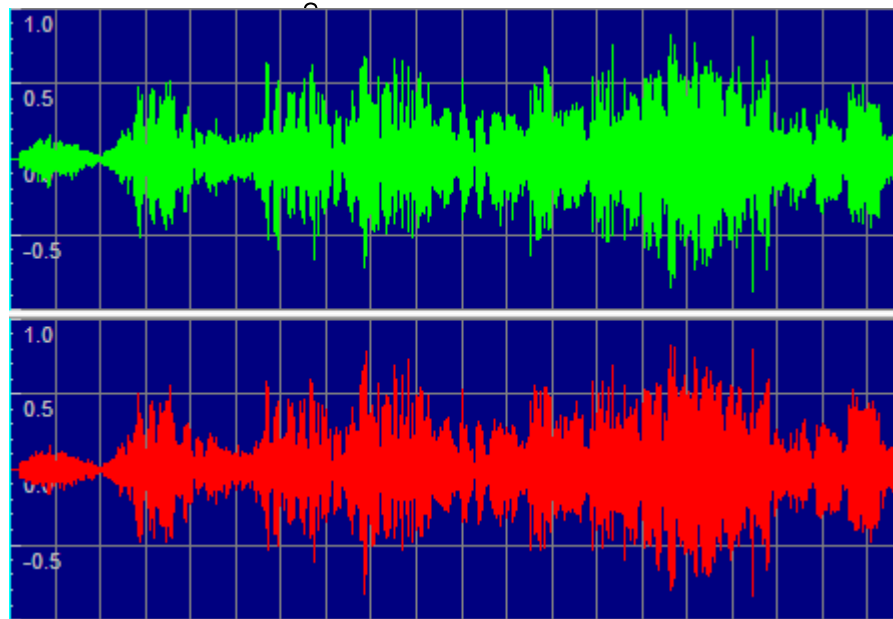
○ 单声道

- 一个声音通道



○ 双声道

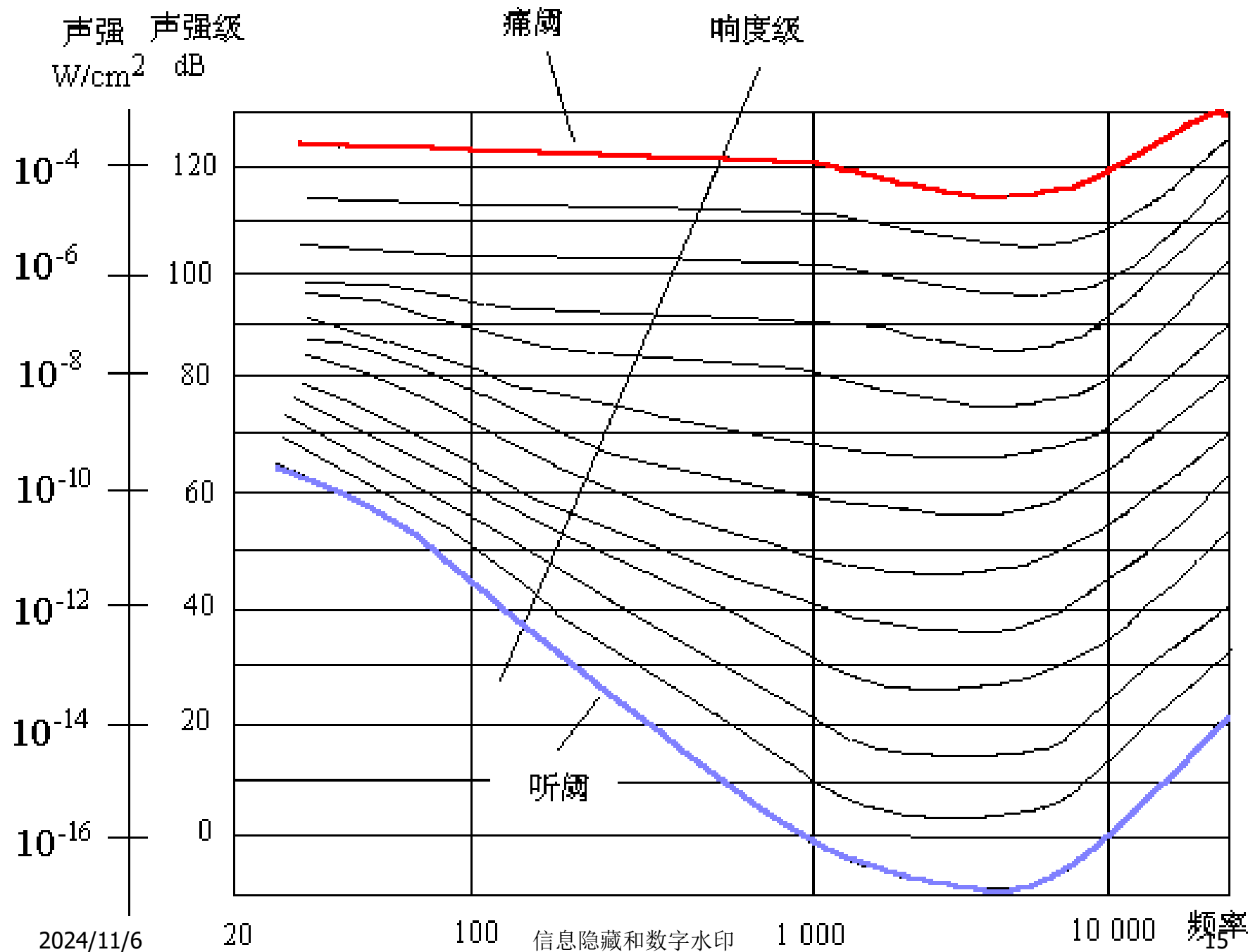
- 二个声音通道，能够形成立体声效果

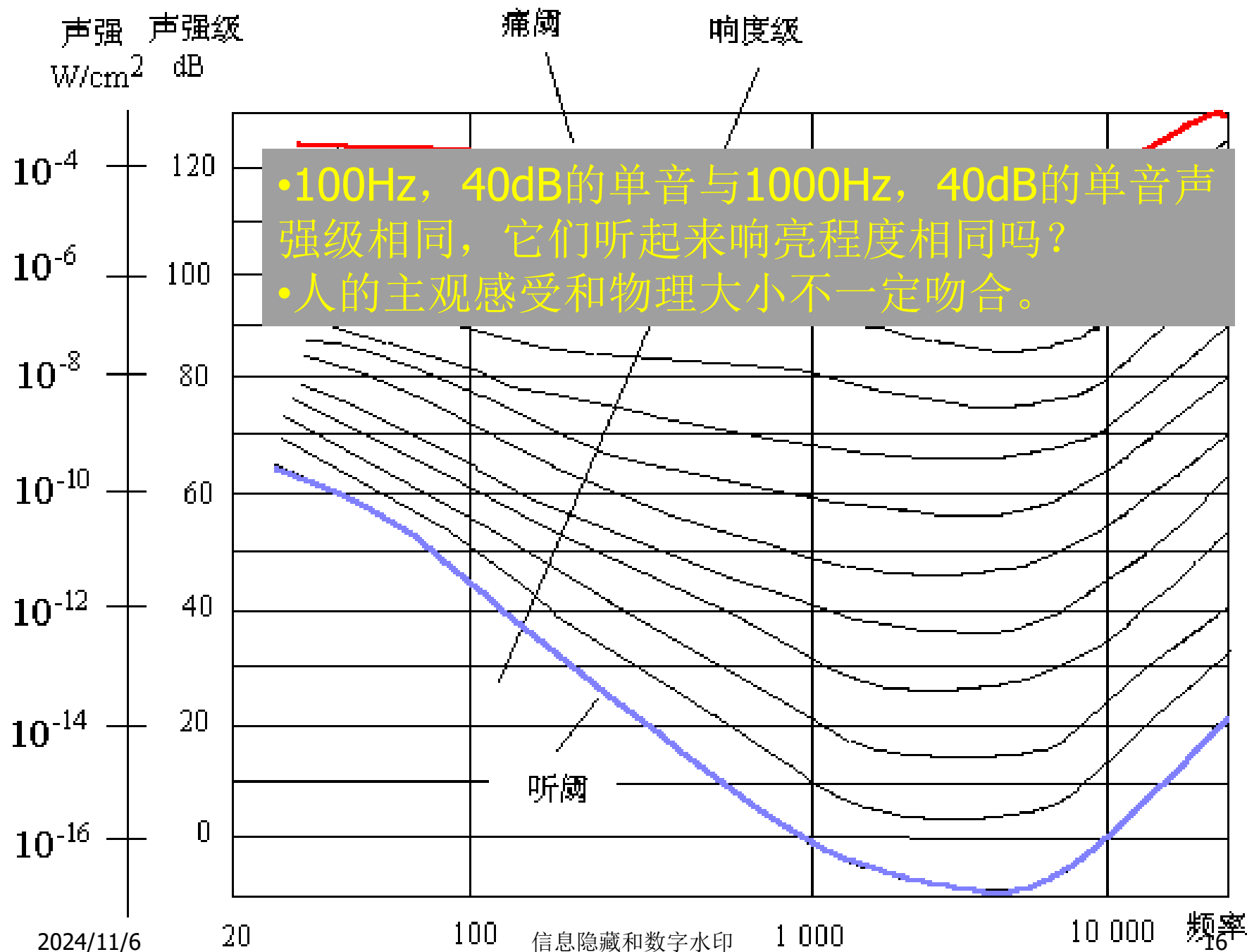


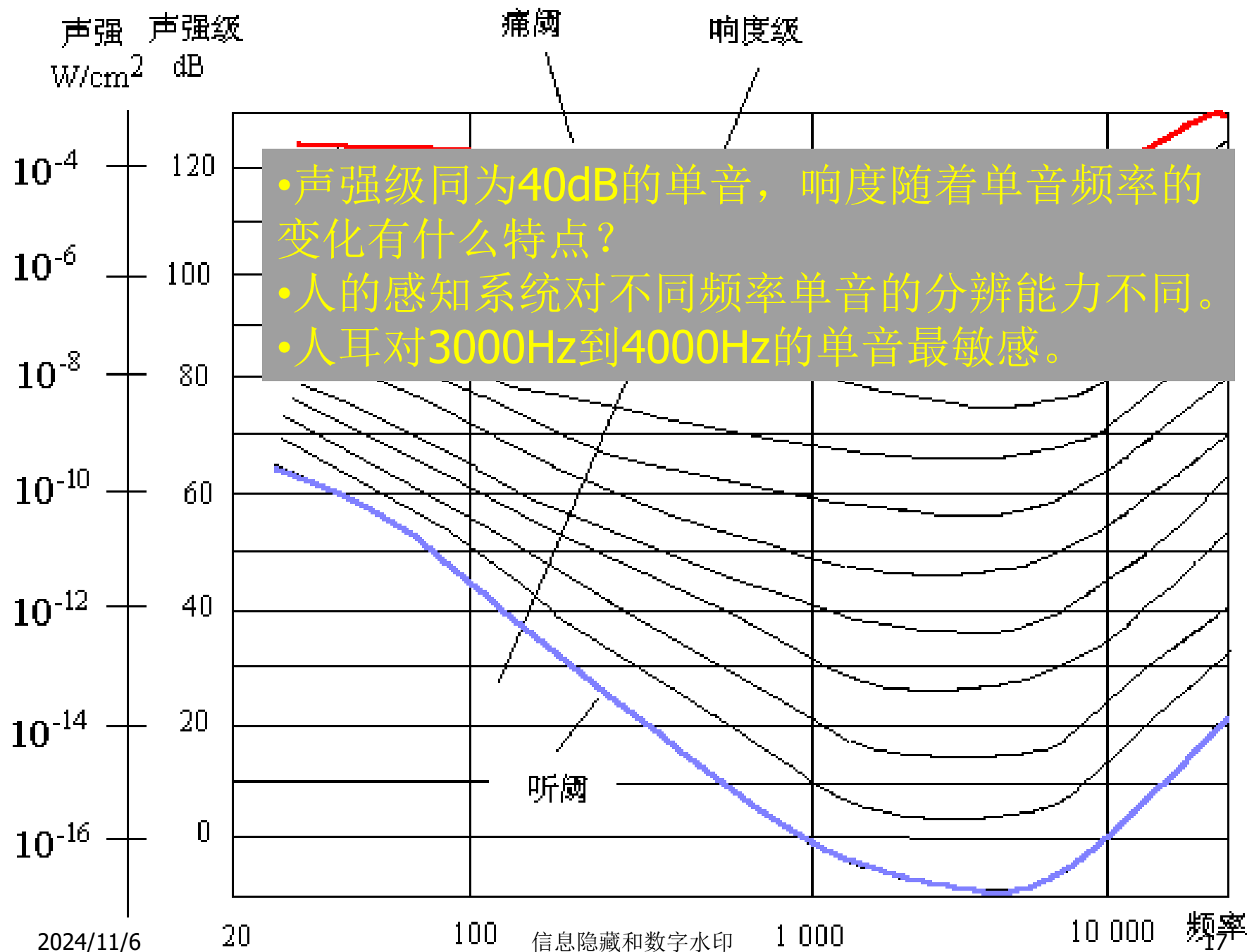
人耳听觉感知特性

——响度

- 响度：人耳对声音强度的主观感受。
- 1方等于1000Hz的1dB纯音的声强级。
- 听阈：0方
- 痛阈：120方







人耳听觉感知特性

- 人类能感知声音的频率范围 **16Hz-16kHz**
- 年轻人, **20kHz**
- 老年人, **10kHz**
- 案例: 公园“驱逐器”

人耳听觉感知特性——掩蔽效应

- 什么是掩蔽效应？
 - 掩蔽效应的“宏观”展示



人耳听觉感知特性——掩蔽效应

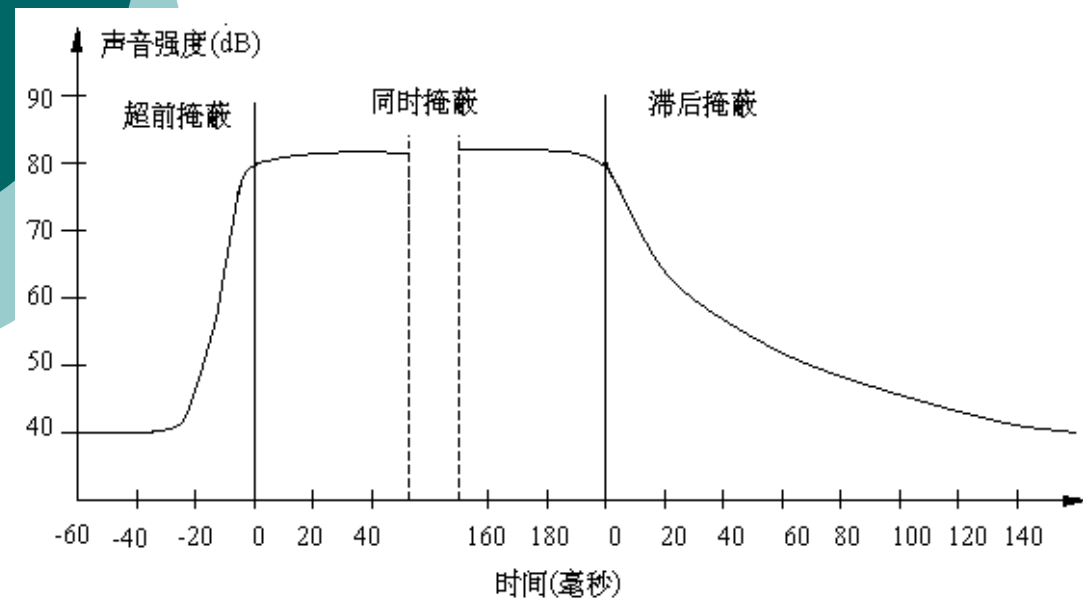


人耳听觉感知特性——掩蔽效应

○ 掩蔽效应

- 心理声学实验表明：人耳难以感知位于强信号附近的弱信号。
- 这种声音心理学现象称为掩蔽。
- 强信号称之为掩蔽音，弱信号称之为被掩蔽音
- 掩蔽音和被掩蔽音同时出现所产生的掩蔽效应称为同时掩蔽或频域掩蔽，
- 否则称之为异时掩蔽或时域掩蔽。

人耳听觉感知特性——掩蔽效应

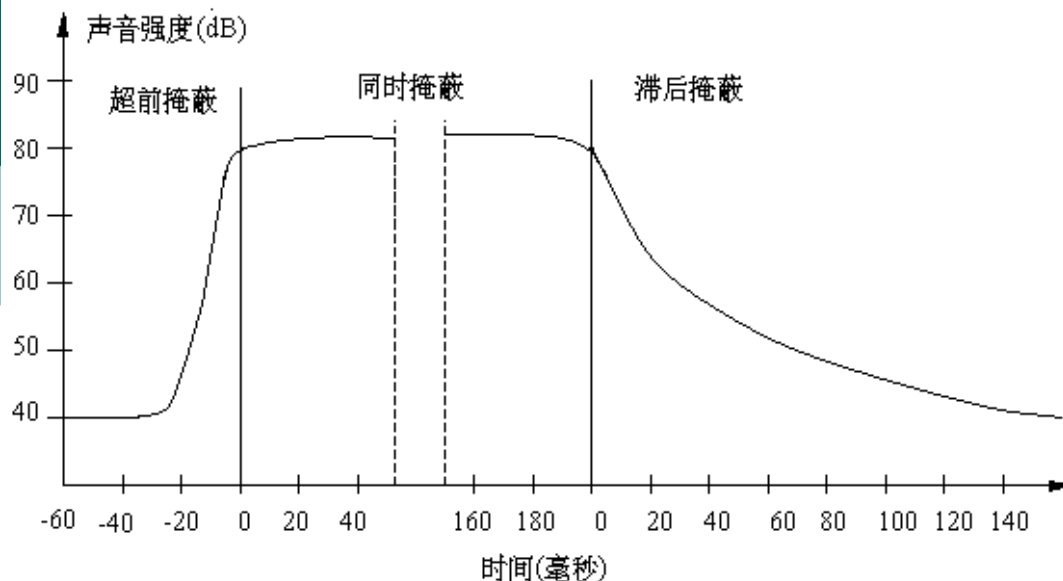


- 同时（频域）掩蔽
 - 低音容易掩蔽高音，而高音掩蔽低音较难。
- 异时（时域）掩蔽
 - 超前掩蔽(pre-masking)
 - 滞后掩蔽(post-masking)

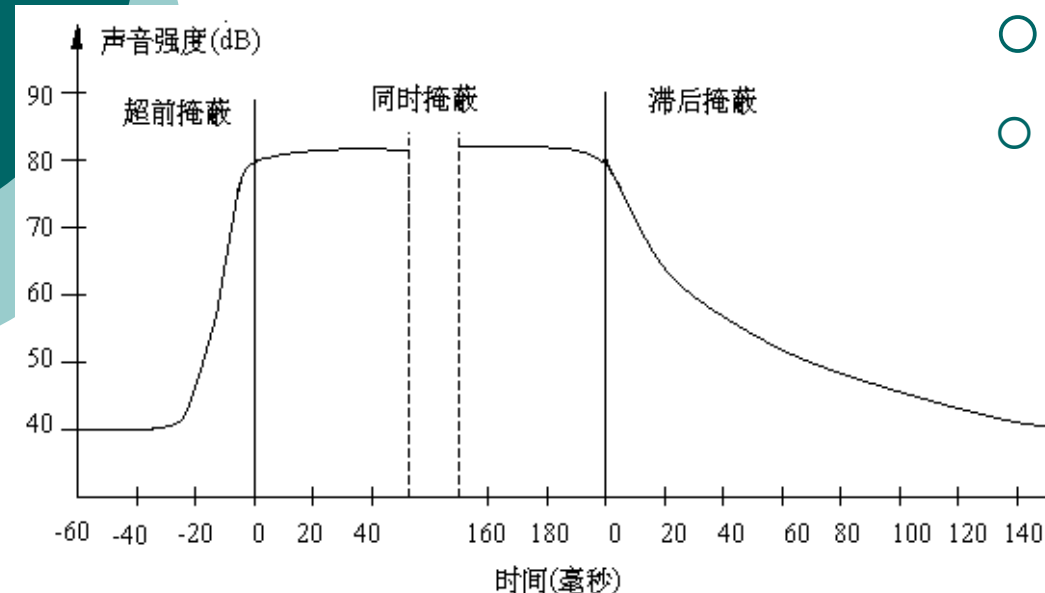
人耳听觉感知特性——掩蔽效应

○ 练习

- 若时刻0时，有一个50dB声强级的单音，20毫秒后，有一个80dB声强级的单音，则人耳能区分出几个声音？
- 若两者出现顺序对调，结果如何？



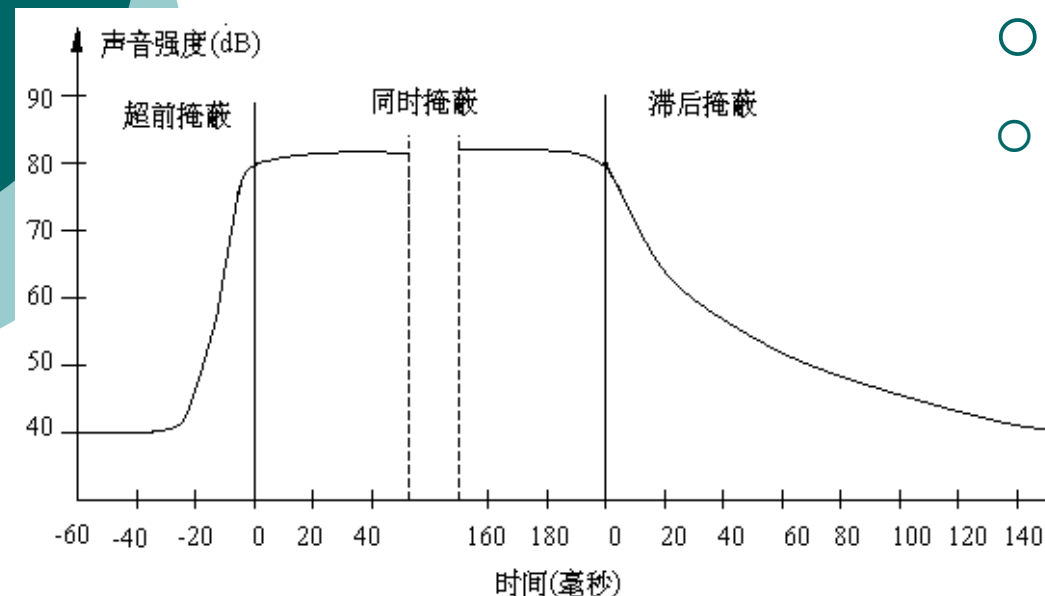
人耳听觉感知特性——掩蔽效应



○ 解答

- 第一问中，弱信号在强信号之前出现，属于超前掩蔽。根据曲线，80dB的强信号，能掩盖在它之前20ms出现的，声强级小于40dB的信号。本例中，弱信号的声强级高于40dB，超出了强信号的掩蔽能力，所以人耳可以分辨出两个信号。

人耳听觉感知特性——掩蔽效应



○ 解答

- 第二问中，弱信号在强信号之后出现，属于滞后掩蔽。根据曲线，80dB的强信号，能掩盖在它之后20ms出现的，声强级小于70dB的信号。本例中，弱信号的声强级小于70dB，在强信号的掩蔽范畴内，所以听起来，只有一个信号存在。

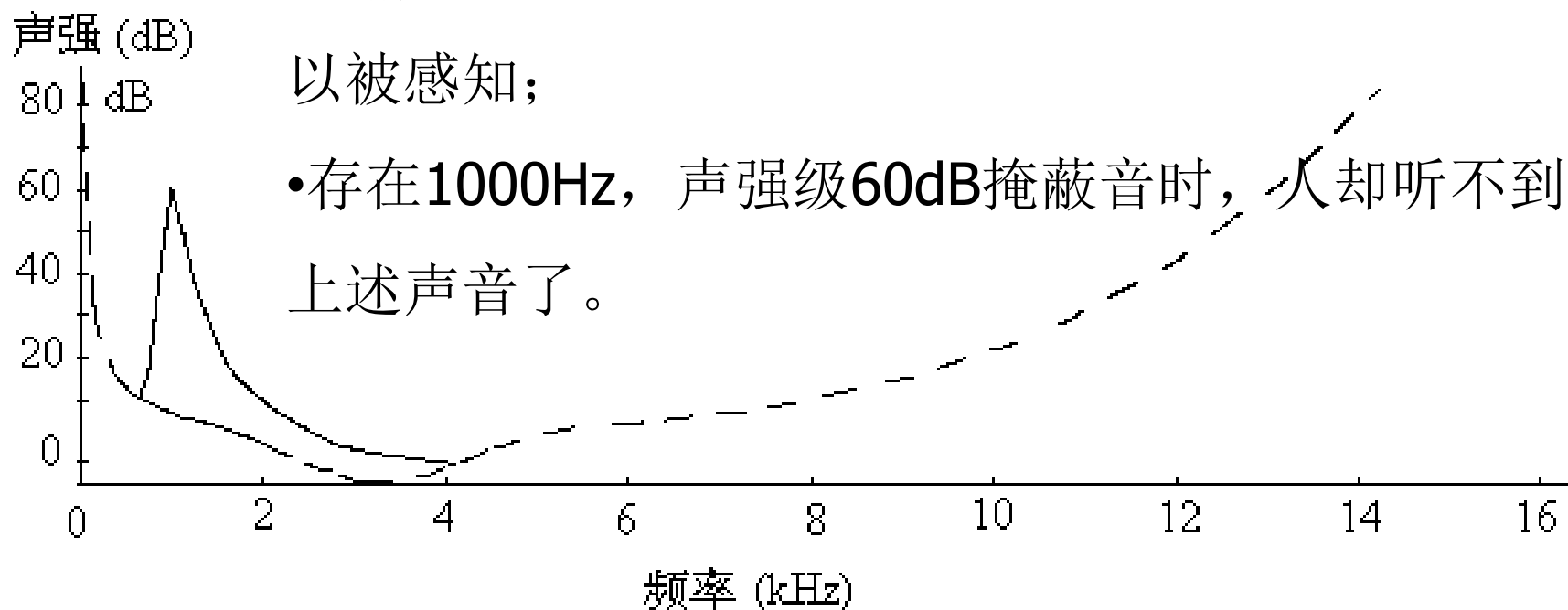
人耳听觉感知特性——掩蔽效应

○ 演示异时掩蔽效应

- **echo_sensible.wav**, 被掩蔽音出现在掩蔽效应持续时间范围外。
- **echo_insensible_bwd.wav**, 被掩蔽音出现在掩蔽效应持续时间范围之内且在掩蔽音之后
- **echo_insensible_fwd.wav**, 被掩蔽音出现在掩蔽效应持续时间范围之内且在掩蔽音之前

人耳听觉感知特性——掩蔽效应

- 虚线所示曲线为听阈。
- 强信号附近的听阈发生了变化——听阈提高。
- 安静环境，2000Hz，声强级5dB左右的声音就可以被感知；
- 存在1000Hz，声强级60dB掩蔽音时，人却听不到上述声音了。



人耳听觉感知特性——掩蔽效应

○ 演示同时掩蔽效应

- 600Hz单音: pich_600Hz.wav
- 1000Hz单音 : pich_1000Hz.wav
- 混合音频:mixer.wav (频域掩蔽)
- 混合音频:mixer_sensible_-20dB.wav (可感知)

语音的质量评价

○ 主观评价

- 主观评价由人对语音的质量进行评价。

○ 客观评价

- 客观评价不以人为主体的。
- 通常提取处理前后的语音的某类特征参量，根据某种规则计算这两组参量之间的差别，是一种“误差”度量。

主观评价方法

- 平均意见分（MOS: Mean Opinion Score）
 - 用5级评分标准来评价语音的质量。

MOS	质量标准	受损程度
5	极好	不可察觉
4	较好	可察觉，但不影响听觉效果
3	一般	轻微影响听觉效果
2	较差	影响听觉效果
1	极差	严重影响听觉效果

- 实验者的分别打分评价语音质量。
- 全体实验者的平均分就是所测语音质量的MOS分。

MOS分评价

○ MOS分

- 4.0-4.5（网络质量）：高质量数字化语音。
- 3.5分左右（通信质量）：能感觉到语音质量有所下降，但不妨碍正常通话。
- 3.0分以下（合成语音质量）：具有足够高的可懂度，但自然度不够好，并且不易进行讲话人识别。

主观评价的优缺点

○ 优点：

- 真实：反映了人对语音质量的真实感觉。

○ 缺点：

- 需要组织大量的试听人员，欠缺灵活性。
- 评分结果重复性和稳定性较低。
- 评分结果受试听者的主观影响较大。

客观评价方法

○ 比较波形的差别:

- 基于**SNR**的评价方法：信噪比、分段信噪比、变频分段信噪比。

$$SNR = 10 \times \lg_{10} \left(\frac{\sum_{i=0}^{N-1} x^2(i)}{\sum_{i=0}^{N-1} (y(i) - x(i))^2} \right)$$

客观评价方法

- 比较参数的差别：
 - 基于LPC的评价方法：比较线性预测编码的系数。
- 比较频谱的差别：
 - 基于谱距离的评价方法。
- 其它评价方法：基于人类听觉模型
- 语音客观评估指标

客观评价方法

- 案例一：语音质量客观指标**SNR**的计算
 - 已知原始音频部分样点值如下：
 - $x[i]: 10, 12, 14, 8, 6, 8$
 - 隐藏信息后，该音频相应样点值变化为：
 - $y[i]: 11, 11, 14, 9, 8, 6$
 - 请计算这部份样点值序列的**SNR**。

客观评价方法

○ 案例一：语音质量客观指标SNR的计算

- 解：首先计算信号能量，根据公式可知：
- $p_s = \sum_{i=0}^5 x^2[i] = 10^2 + 12^2 + \dots + 8^2 = 604$
- 再计算噪声能量，原始音频和水印音频的差值即为噪声：
- $p_n = \sum_{i=0}^5 n^2[i] = \sum_{i=0}^5 (y[i] - x[i])^2 = (11 - 10)^2 + (11 - 12)^2 + \dots + (6 - 8)^2 = 11$
- 最后计算信号与噪声的比值
- $snr = p_s / p_n = 604 / 11 \approx 54.91$
- 换算为dB: $snr = 10 \lg(54.91) \approx 17.40 \text{dB}$

客观评价方法

- 演示

- 计算A律压扩引入的信噪比

$$SNR = 10 \times \lg_{10} \left(\frac{\sum_{i=0}^{N-1} x^2(i)}{\sum_{i=0}^{N-1} (y(i) - x(i))^2} \right)$$

略过深入讨论

客观评价与主观评价一致性的问题

- 研究方便、快捷、准确的，与主观评价相一致的客观评价标准，是该领域研究的一个方向。
- 如何衡量一个客观评价与主观评价的一致性？

客观评价与主观评价一致性的问题

- 通过在客观评价和主观评价之间建立的函数关系，可以用客观评价值求出对主观评价值的预测值。
- 该预测值和实测的主观评价值之间的相关度就作为该客观评价方法与主观评价方法之间的相关度。

客观评价与主观评价一致性的问题

$$\rho = \sqrt{\frac{\sum_{i=1}^N (\hat{S}_i - \mu)^2}{\sum_{i=1}^N (S_i - \mu)^2}}$$

- N为被测的样本数， s_i 表示第i个样本的实测主观评价值， \hat{s}_i 表示第i个样本的客观评价的主观预测值， μ 是实测主观评价值的算术平均值
- ρ 值越接近1，说明该客观评价方法对主观评价的预测越准确，该方法的性能越好

客观评价的优缺点

○ 优点：

- 简单，可重复性强，与试听者主观感觉无关。

○ 缺点：

- 无法达到与主观评价完全一致的效果。

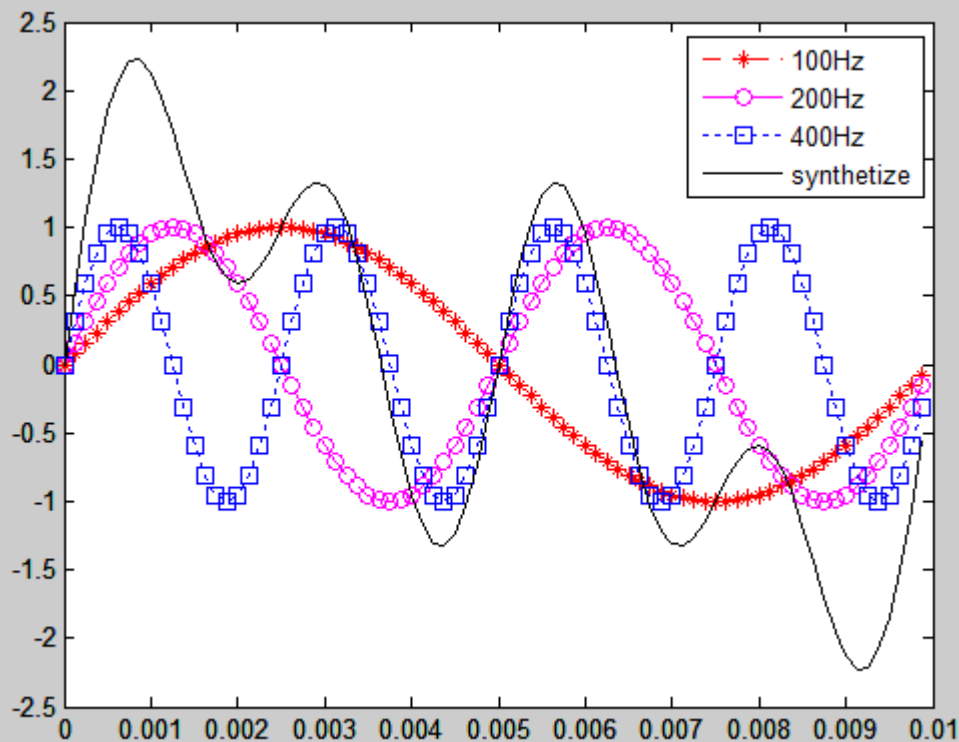
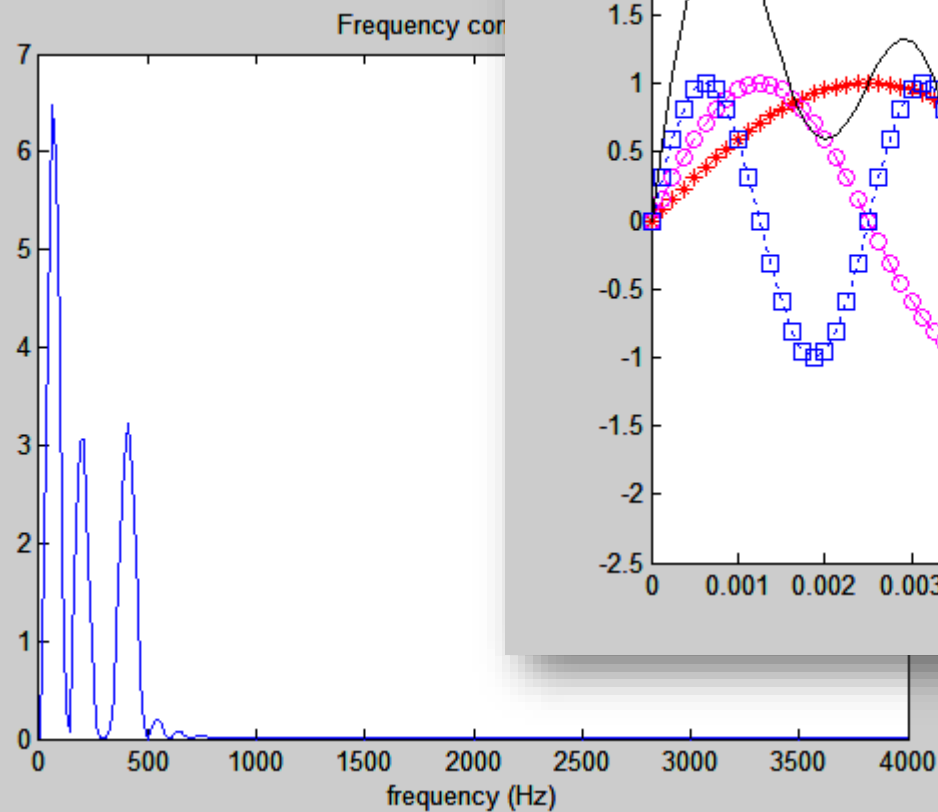
常用语音信号变换

○ 常用变换域操作

- DFT(Discrete Fourier Transform): 离散傅里叶变换
- DCT(Discrete Cosine Transform): 离散余弦变换
- DWT(Discrete Wavelet Transform): 离散小波变换

常用语音信号变换

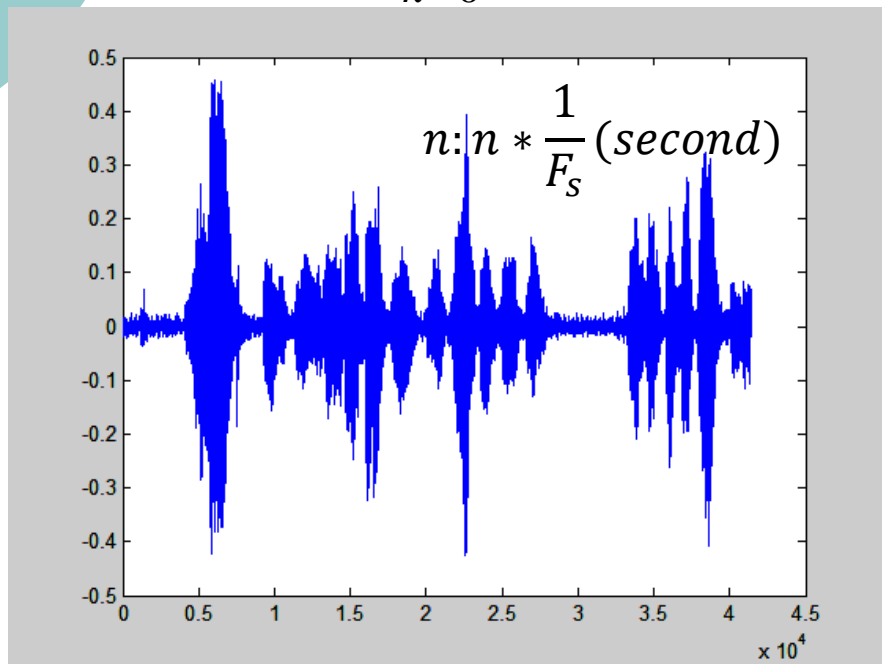
○ 用傅里叶



常用语音信号变换

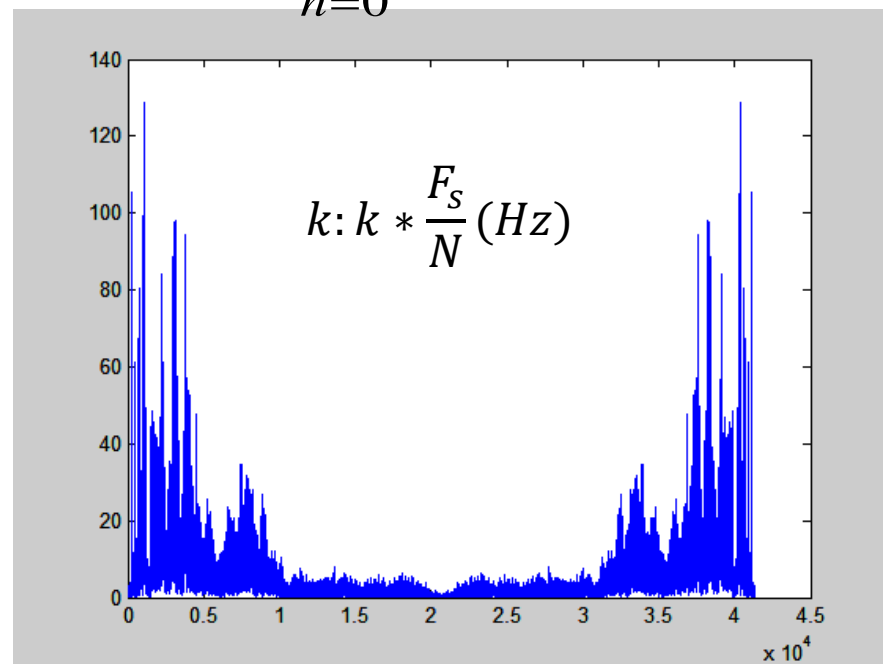
音频（时域）

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}$$



音频（**DFT**）

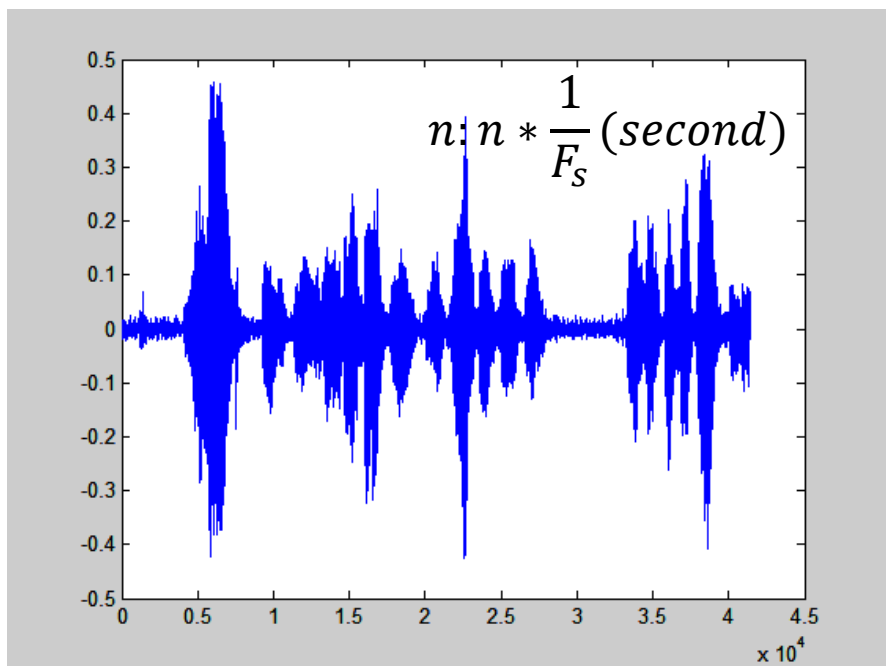
$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}$$



常用语音信号变换

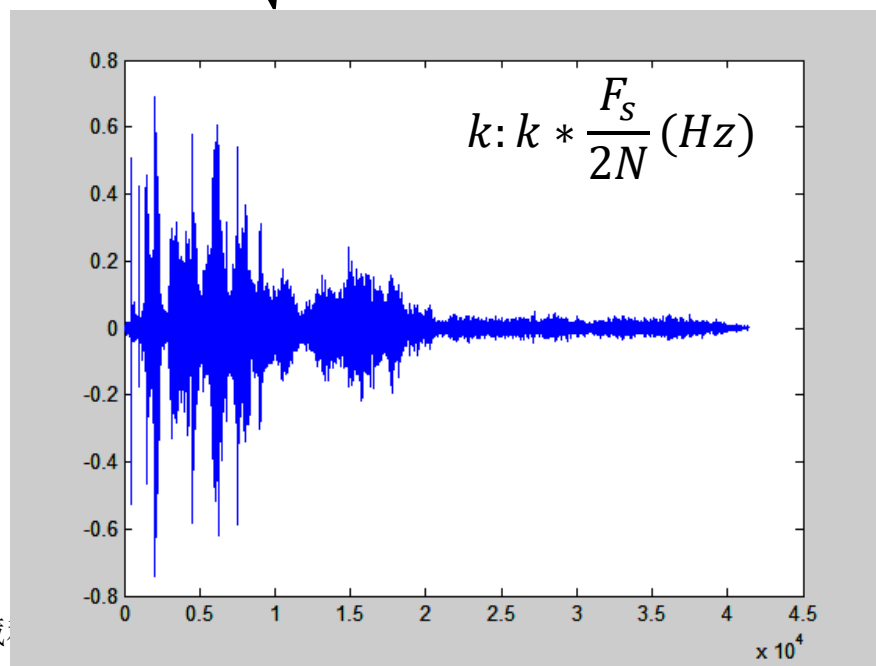
音频（时域）

- $x[n] = \frac{1}{\sqrt{N}} X[0] + \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} X[k] \cos \frac{(2n+1)k\pi}{2N}$



音频（**DCT**）

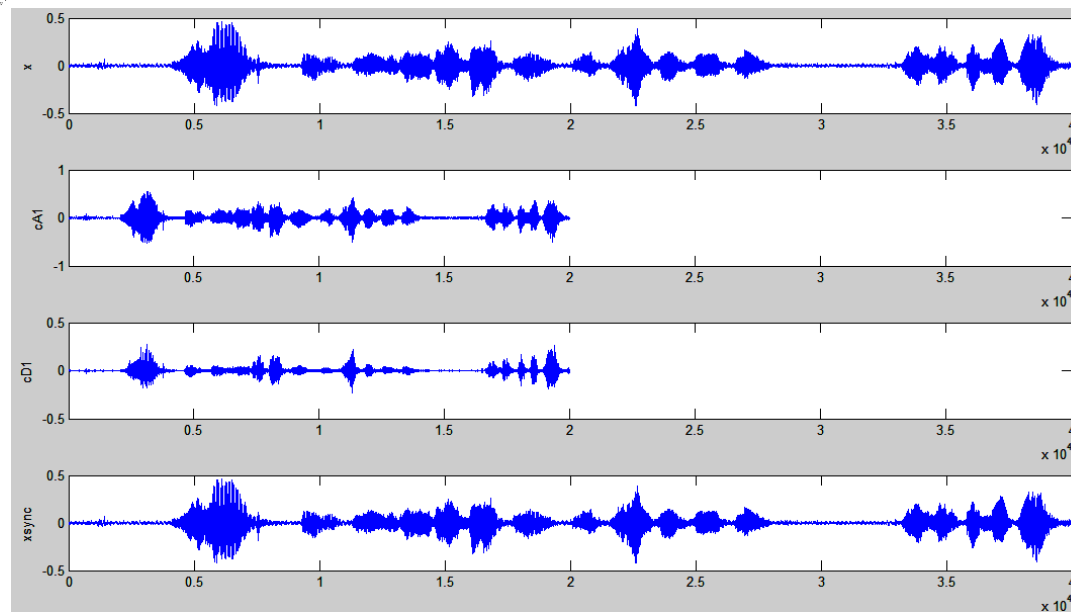
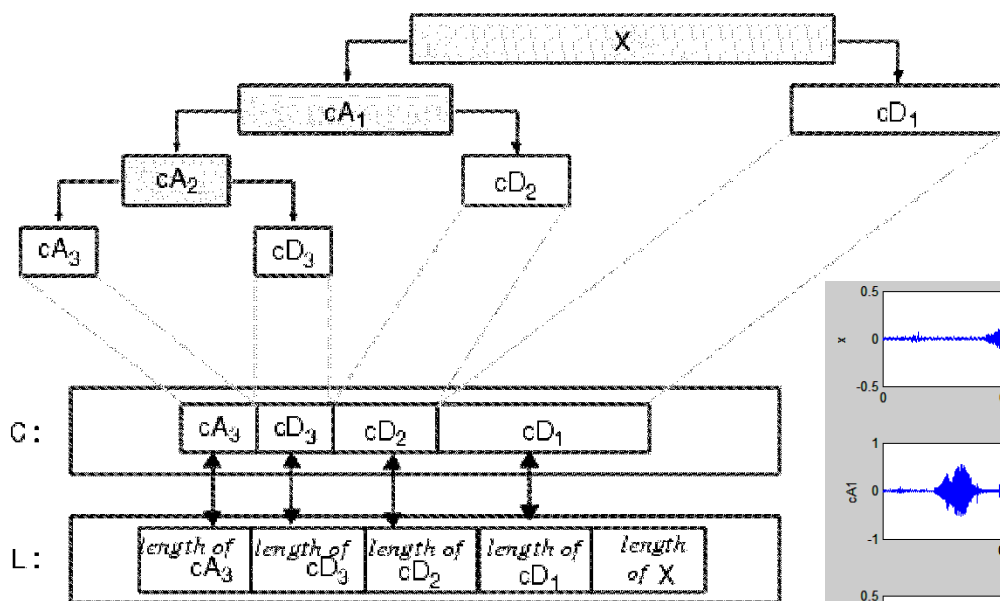
- $X[0] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n]$
- $X[k] = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x[n] \cos \frac{(2n+1)k\pi}{2N}$



常用语音信号变换

DWT原理和系数

音频 (DWT)



常用语音信号变换

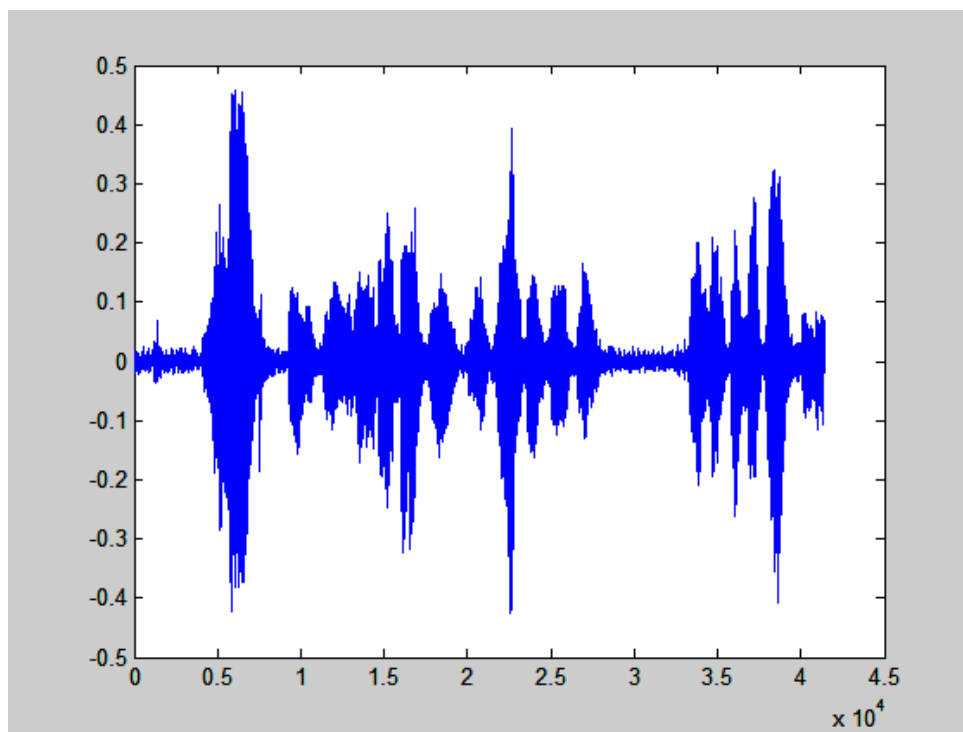
○ 案例二：DWT系数特性研究

- DWT可迭代地对信号进行多层次分解，其中近似系数反映信号的平缓变换情况，可视为信号低频成份，细节系数反映信号剧烈变换情况，可视为信号高频成份。
- 案例尝试修改语音信号DWT细节系数，观察修改前后语音质量变换情况。

常用语音信号变换

○ 案例二：DWT系数特性研究

● 1. 准备语音信号

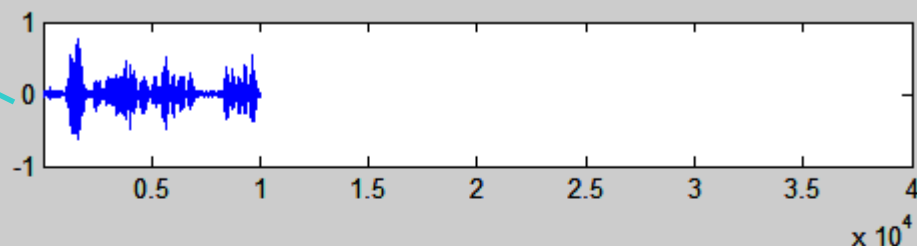


常用语音信号变换

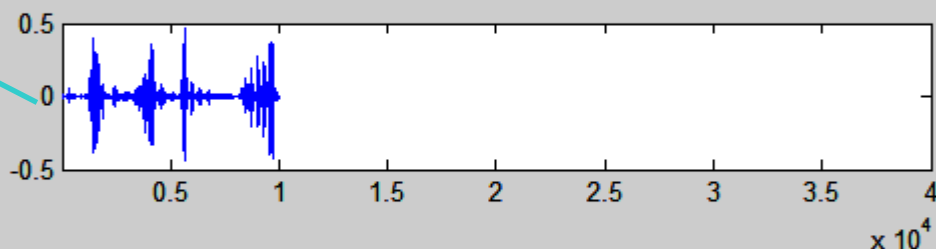
○ 案例二：DWT系数特性研究

- 2. 对语音信号进行2级小波分解，得到各级小波系数。

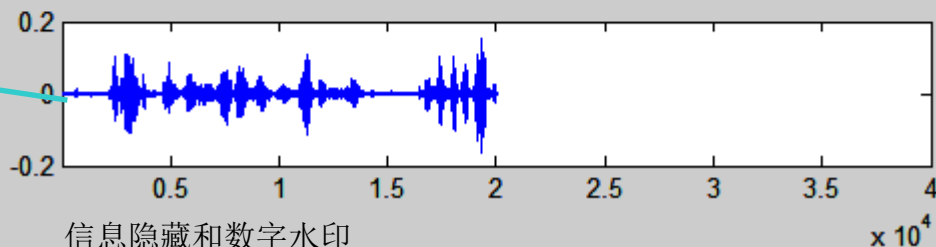
2级近似系数



2级细节系数



1级细节系数



常用语音信号变换

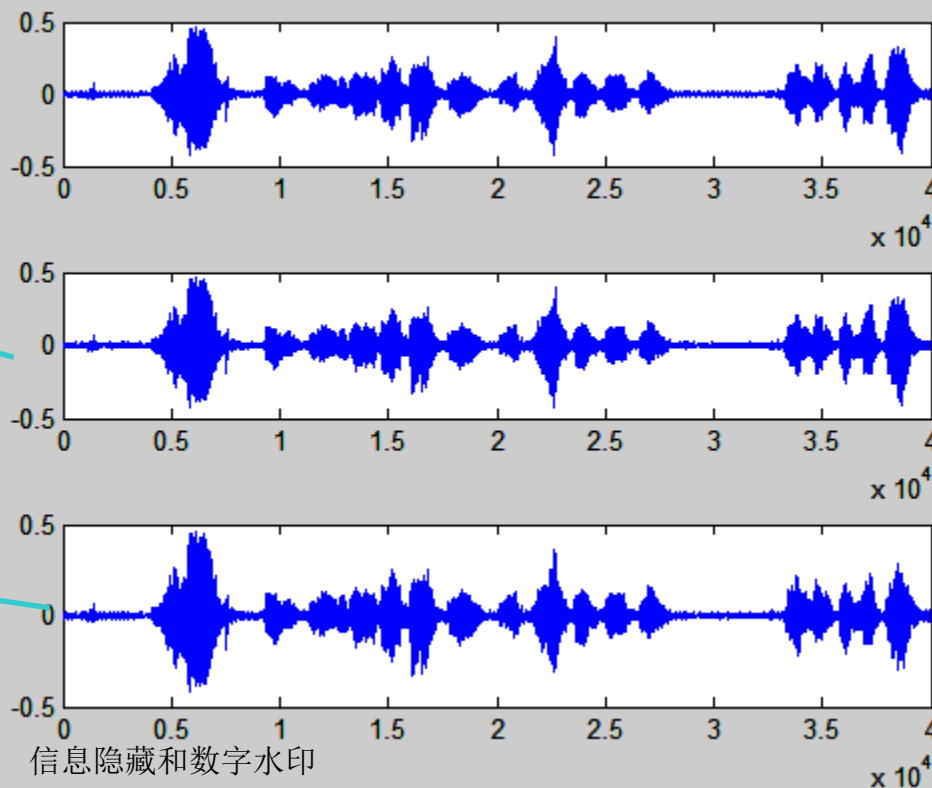
案例二：DWT系数特性研究

- 3. 分别用两组系数重构语音。第一组：原始系数，第二组：一半2级细节系数置零。

原始音频

原始系数重构音频

修改系数重构音频



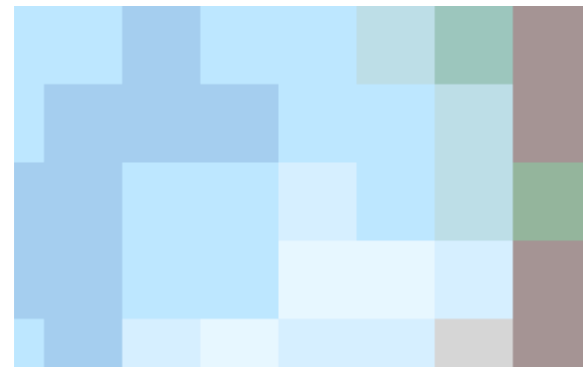
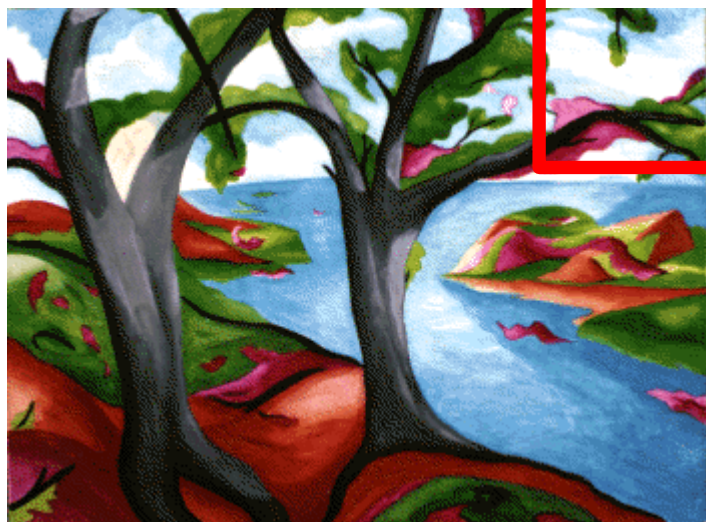
图像信号处理基础

- 数字图像的概念
- 人类视觉系统(HVS)特点
- 图像质量评价
- 图像信号处理基础

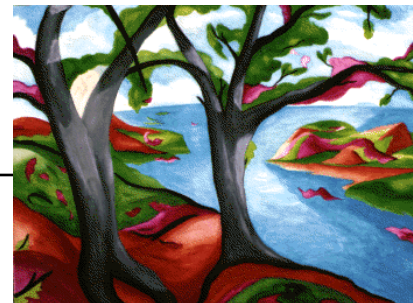
图像信号处理基础

○ 图像的基本表示

- 图像由像素（**pixel**）组成，像素存储为矩阵。
- $I=f(x,y), 0 \leq x \leq M-1, 0 \leq y \leq N-1$



图像信号处理基础



○ 图像分辨率（image resolution）

- 图像像素数量给出了像素总数的信息，图像分辨率给出像素点分布的密度，常记为：水平*垂直，分别表示水平和垂直方向上，单位打印长度上分布的像素数量。
- 例如：显示屏分辨率通常为**1024*768**，表示水平方向，单位长度上有**1024**个像素；垂直方向，单位长度上有**768**个像素。
- 示例图像为**350*258**

图像信号处理基础



- 常见图像类型包括
 - 灰度图像($I=f(x,y), 0 \leq I \leq 255$)
 - 像素表示图像在对应网格位置的明暗程度，即像素矩阵存储图像灰度级信息。
 - 从纯黑到纯白，总共有**256**个灰度级。其中纯黑的灰度级为**0**，纯白的灰度级为**255**。
 - 每个像素用**1**个字节表示。

170	170	169	170	169	166	166	166
170	170	169	170	169	166	166	166
170	170	169	170	169	166	166	166
170	170	169	170	169	166	166	166
170	170	169	170	169	166	166	166
170	170	168	166	168	166	165	166
169	169	168	165	166	168	167	165
166	166	166	165	165	166	166	164

图像信号处理基础

○ 常见图像类型包括

● 24位彩色图像

- 每个像素用三个字节表示，其含义为该像素颜色的红、绿、蓝分量。

- $I=f(x,y,c), 0 \leq c \leq 2,$

$$0 \leq I \leq 255,$$

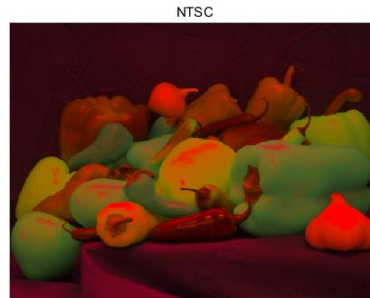
$$M \times N \times C$$

第1通道, $f(:, :, 0)$, 为红色分量



62	63	63	65	66	63	61	63			
63	61	59	64	63	60	61	64			
65	29	31	34	30	27	31	31	30		
63	31	31	32	30	28	31	31	32		
63	29	30	64	64	64	60	59	62	62	61
63	29	29	62	64	64	60	59	61	62	62
62	31	32	60	62	63	61	61	64	63	63
65	32	32	62	63	63	60	62	62	62	62
	30	30	62	63	62	61	61	60	60	61
	29	29	61	61	62	59	59	61	65	62
			63	61	60	59	61	63	66	62
			65	60	59	60	62	64	63	61

图像信



NTSC彩色空间

亮度（Y）、色调（I）和饱和度（Q）；分离灰度信息和颜色信息

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

YCbCr彩色空间

亮度（Y）、色差分量Cb（蓝色分量和参考值的差）和色差分量Cr（红色分量和参考值的差）

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

更多彩色空间

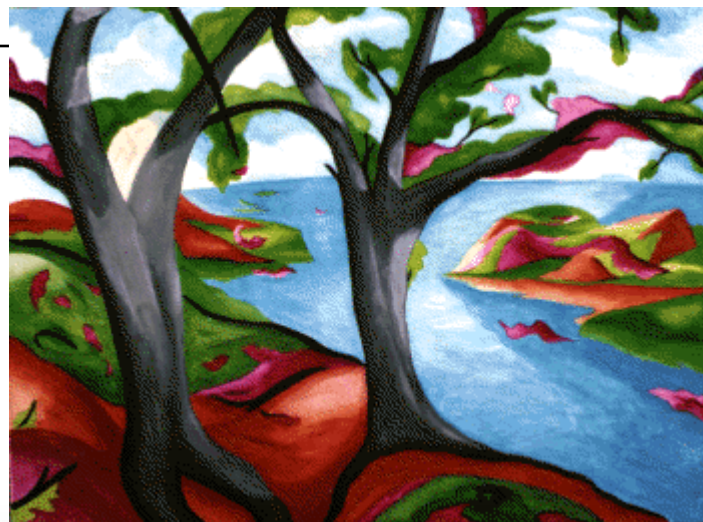
HSV（色调、饱和度、数值），更接近人们对色彩的感知。在画家术语里，色调、饱和度、数值被称为色泽、明暗和调色。

CMY（青色、品红色和黄色），CMYK（青色、品红色、黄色和黑色）。

HSI（hue，色度; Saturation，饱和度; intensity，饱和度）

图像信号处理基础

- 常见图像类型包括
 - 索引（调色板）图
 - 像素矩阵中每个元素存储一个调色板索引。
 - 调色板每条记录表示一种颜色。



108	116	98	108	87	90	90	86
112	112	112	105	112	105	112	105
112	112	105	0	0	0		
105	105	105	0.0627	0.0627	0.0314		
112	105	112	0.2902	0.0314	0		
105	105	105	0	0	1.0000		
105	112	105	0.2902	0.0627	0.0627		
105	105	112	0.3882	0.0314	0.0941		
			0.4510	0.0627	0		
			0.2588	0.1608	0.0627		

图像信号处理基础

- 常见图像类型包括
 - 黑白（二值）图像
 - 像素矩阵元素值为0或1，
 - 对应颜色为黑色，白色



1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

图像信号处理基础



人类视觉特点

○ 视觉范围

- 视觉范围是指人眼所能感觉的亮度范围。
- 客观上相同的亮度，当平均亮度不同时，主观感觉的亮度也不相同。

人类视觉特点

○ 视觉范围

- 视觉范围非常宽，但是人眼并不能同时感受这样宽的亮度范围。
- 当平均亮度比较适中时，能分辨的亮度的范围较大。
- 当平均亮度较低时，能分辨的亮度范围较小。

人类视觉特点

○ 视觉适应性

● 暗适应性

- 人由亮处走到暗处时的视觉适应过程，称为暗适应。
- 暗适应过程约需大约三十分钟。

● 亮适应性

- 人由暗处走到亮处时的视觉适应过程，称为亮适应。
- 亮适应过程约需大约一分钟。

人类视觉特点

○ 视觉惰性（视觉暂留）

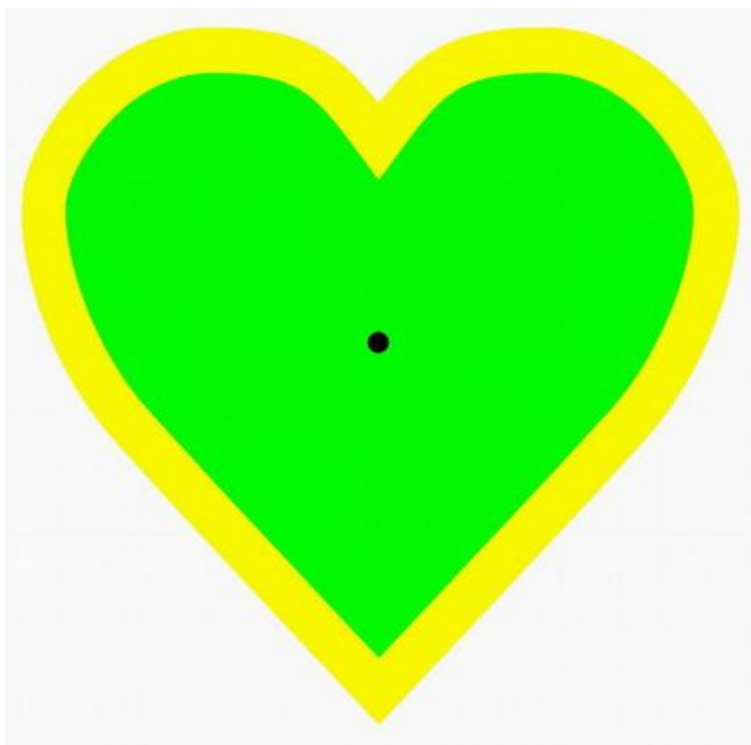
- 人眼对亮度改变进行跟踪的滞后性质称为视觉惰性。
- 人眼所看到的影像消失后，人眼仍能继续保留其影像约0.1-0.5秒左右。

人类视觉特点

○ 视觉惰性（视觉暂留）

- 看纹理丰富的窗帘，然后再看墙壁，看墙时会发现窗帘的花纹重叠在墙上了。
- 看电视，电影是每分钟**20**几帧图片，通过暂留形成了流畅的动作。

人类视觉特点



(图片源自参考文献3)

- 视觉惰性（视觉暂留）
 - 持续注视图案中心黑点30秒钟以上。
 - 迅速看向无字白纸或空白墙壁。
 - 将看到视觉暂留。

图像质量评价——主观评价方法

○ 主观评价

● 平均意见分

- MOS (Mean Opinion Score)
- 观察者根据一些事先规定的评价尺度,
- 或自己的经验,
- 判断测试图像的视觉效果质量,
- 所有观察者给出的分数的加权平均。
- 五级评分

图像质量评价——主观评价方法

○ 主观评价

在ImageNet上，可以签订协议后展开实验。提供一对图像给用户，包含原图和失真图，让用户选择更自然/真实的图像。如果正确率在50%，那么说明人眼完全无法识别失真。

4.3 HUMAN PERCEPTUAL STUDY

To quantify the perceptual realism of stAdv's adversarial examples, we perform a user study with human participants on Amazon Mechanical Turk (AMT). We follow the same perceptual study protocol used in prior image synthesis work (Zhang et al., 2016; Isola et al., 2017). We generate 600 images from an ImageNet-compatible dataset, described in Appendix C. In our study, the participants are asked to choose the more *visually realistic* image between an adversarial example generated by stAdv and its original image. During each trial, these two images appear side-by-side for 2 seconds. After the images disappear, our participants are given unlimited time to make their decision. To avoid labeling bias, we allow each user to conduct at most 50 trails. For each pair of an original image and its adversarial example, we collect about 5 annotations from different users.

In total, we collected 2,740 annotations from 93 AMT users,. Examples generated by our method were chosen as the more realistic in $47.01\% \pm 1.96\%$ of the trails (perfectly realistic results would achieve 50%). This indicates that our adversarial examples are almost indistinguishable from natural images.

图像质量评价——主观评价方法

- 主观评价的优点：
 - 比较准确。
- 主观评价的缺点：
 - 受观察者的知识背景、情绪，
 - 和疲劳程度等因素的影响。
 - 可重复性较差。

图像质量评价——客观评价方法

○ 客观评价

- 处理和分析系统中输入和输出的图像信号，
- 提取图像特征参量作为研究分析对象，
- 比较特征参量，以此反映图像间的差别 。

图像质量评价——客观评价方法

○ 常用客观评价方法

● 均方误差（MSE, Mean Square Error）

- $$MSE = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x[i, j] - y[i, j])^2$$

● 平均绝对误差（MAE, Mean Absolute Error）

- $$MAE = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} |x[i, j] - y[i, j]|$$

图像质量评价——客观评价方法

○ 常用客观评价方法

- 峰值信噪比（PSNR, Peak Signal to Noise Ratio）

- $$PSNR = 10lg \left(\frac{x_{Max}^2[i,j]}{MSE} \right)$$

图像质量评价——客观评价方法

SSIM（结构相似性指数，Structural Similarity Index Measure）

平均灰度作
为亮度测量

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i.$$

灰度标准差作
为对比度测量

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}}.$$

协方差作为结构测量

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y).$$

亮度对比函数

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

对比度对比函数

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

结构对比函数

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}.$$

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma$$

图像质量评价——客观评价方法



噪声模式：高斯噪声 PSNR:29.9630dB SSIM:0.9819

图像质量评价——客观评价方法

- 案例三：图像质量客观指标PSNR的计算
 - 已知原始图像部分像素值如下：
 - $x[i,j]: 10, 12, 14,$
 - $8, 6, 8$
 - 隐藏信息后，该图像相应像素值变化为：
 - $y[i,j]: 11, 11, 14,$
 - $9, 8, 6$
 - 请计算这部份像素构成的子图的PSNR。

图像质量评价——客观评价方法

- 案例三：图像质量客观指标PSNR的计算
 - 解：首先计算均方误差，MSE，由公式得：
 - $MSE = \frac{1}{2 \times 3} \sum_{i=0}^1 \sum_{j=0}^2 (x[i,j] - y[i,j])^2 = \frac{1}{6} \times [(11 - 12)^2 + \dots + (6 - 8)^2] \approx 1.83$
 - 再计算PSNR，以灰度图考虑，最大像素值为255，则：
 - $PSNR = 10 \lg \left(\frac{x_{Max}^2[i,j]}{MSE} \right) = 10 \lg \left(\frac{255^2}{1.83} \right) \approx 45.50dB$

图像质量评价——客观评价方法

○ 客观评价优点：

- 使用方便。
- 可重复性强。

○ 客观评价缺点：

- 评价结果与主观感觉有时存在不一致。

图像的质量评价



- 为什么客观评价与主观评价不能很好吻合？
 - 人眼对所看到物体的理解，
 - 不仅与生理因素有关，
 - 还在相当大的程度上取决于心理因素，
 - 如“视而不见”，“听而不闻”。

图像的质量评价



- 为什么客观评价与主观评价不能很好吻合？
 - 对感兴趣的区域给予极大关注，对其它区域不在意。
 - 大脑对所接收的事务有一个过滤和取舍的过程，目前计算机还无法很好地模拟此过程。

例1



- 高斯噪声，椒盐噪声，PSNR相同

	
PSNR(dB)	PSNR(dB)
48.1183	50.9538



例1（续）

	
PSNR(dB)	PSNR(dB)
34.1389	35.1996

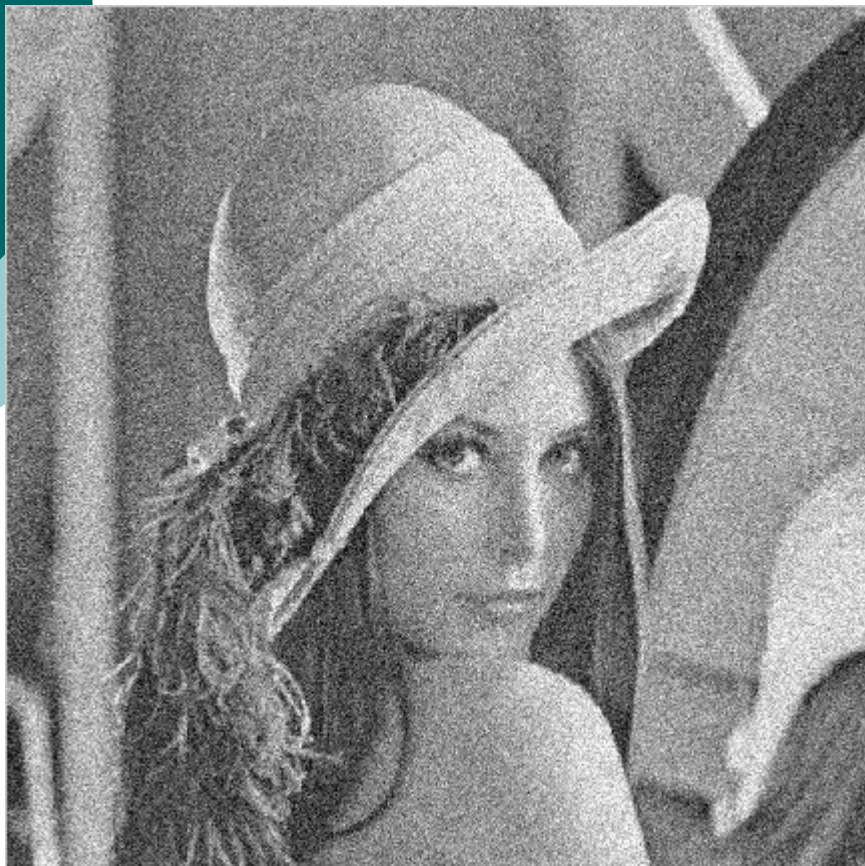
例1（续）

	
PSNR(dB)	PSNR(dB)
28.1183	29.0575

例1（续）

	
PSNR(dB)	PSNR(dB)
19.5175	19.0568

例1（续）





高斯噪声（PSNR \approx 19.5dB）

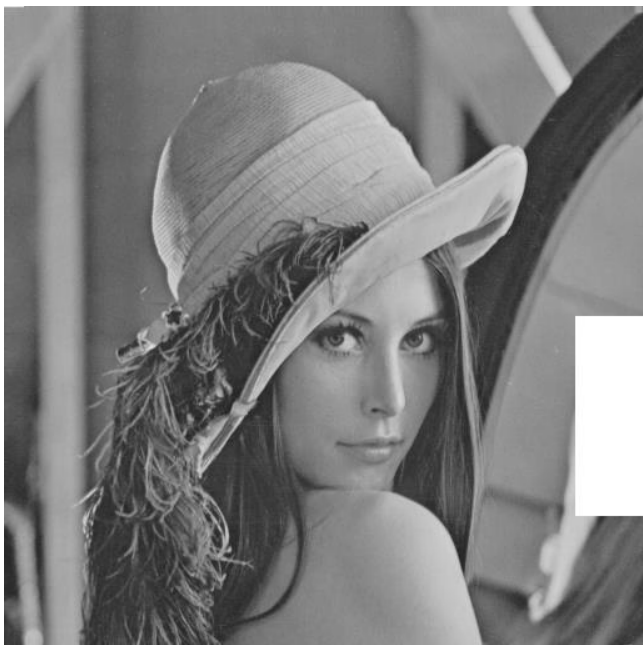



椒盐噪声（PSNR \approx 19.0dB）



例2

	
PSNR(dB)	PSNR(dB)
23.1805	23.5164

例2（续）

	
PSNR(dB)	PSNR(dB)
23. 5084	23. 2079

例2（续）

	
PSNR(dB)	PSNR(dB)
33.8497	33.5907

常用图像信号变换

○ 常见变换

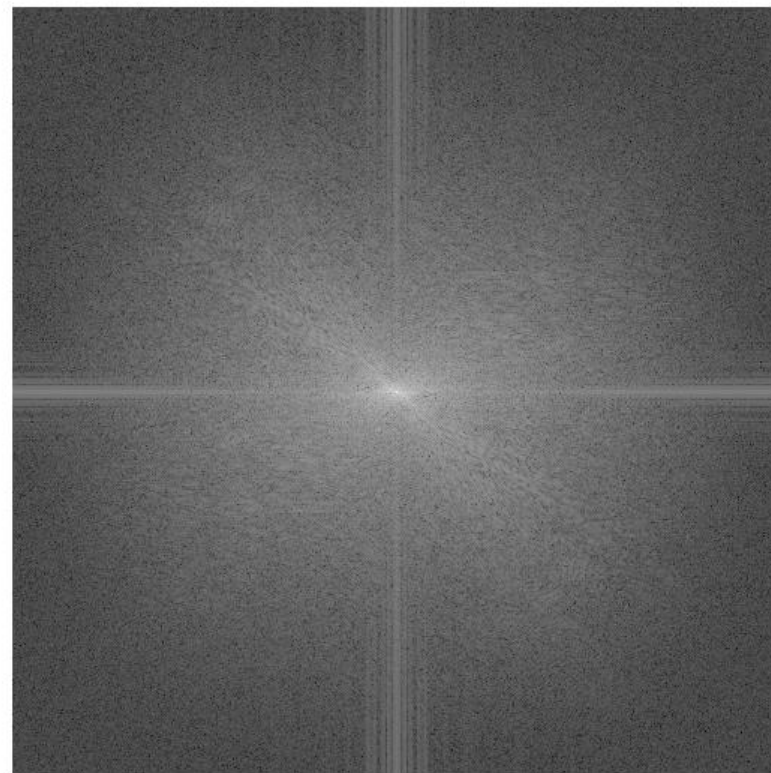
- DFT(Discrete Fourier Transform): 离散傅里叶变换
- DCT(Discrete Cosine Transform): 离散余弦变换
- DWT(Discrete Wavelet Transform): 离散小波变换

常用图像信号变换

图像（空域）



图像（**DFT**）



常用图像信号变换

图像（空域）



图像（**DCT**）



常用图像信号变换

图像（空域）



图像（**DWT**）

cA2	cH2	cH1
cV2	cD2	
cV1		cD1

常用图像信号变换

图像（空域）



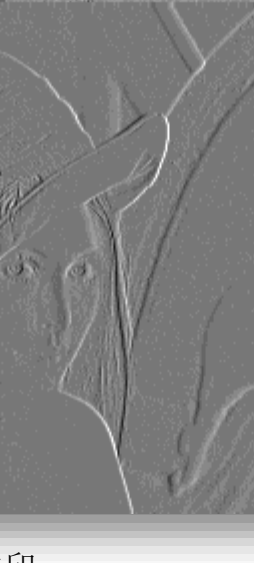
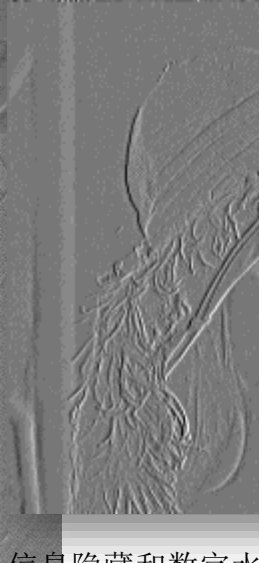
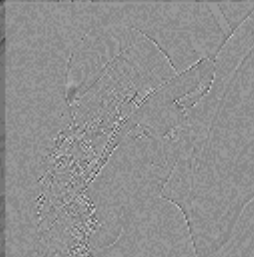
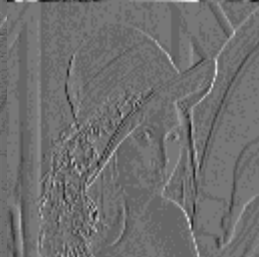
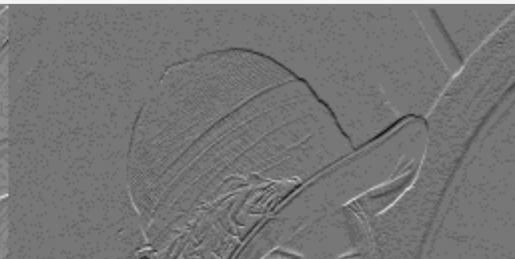
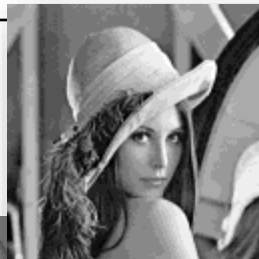
图像（**1级DWT**）



常用图像信号变换

2级DWT

1级DWT



常用图像信号变换

○ 案例四：DCT系数特性研究

- DCT常用于图像信号处理操作，例如，JPEG图像压缩。图像不同位置的DCT系数大小，反映图像不同频率成份的强弱。位置与频率有对应关系，水平方向从左到右，频率逐渐升高；垂直方向从上到下，频率逐渐升高。
- 案例尝试修改图像信号DCT高频部份系数，观察修改前后图像质量变换情况。

常用图像信号变换

○ 案例四：DCT系数特性研究

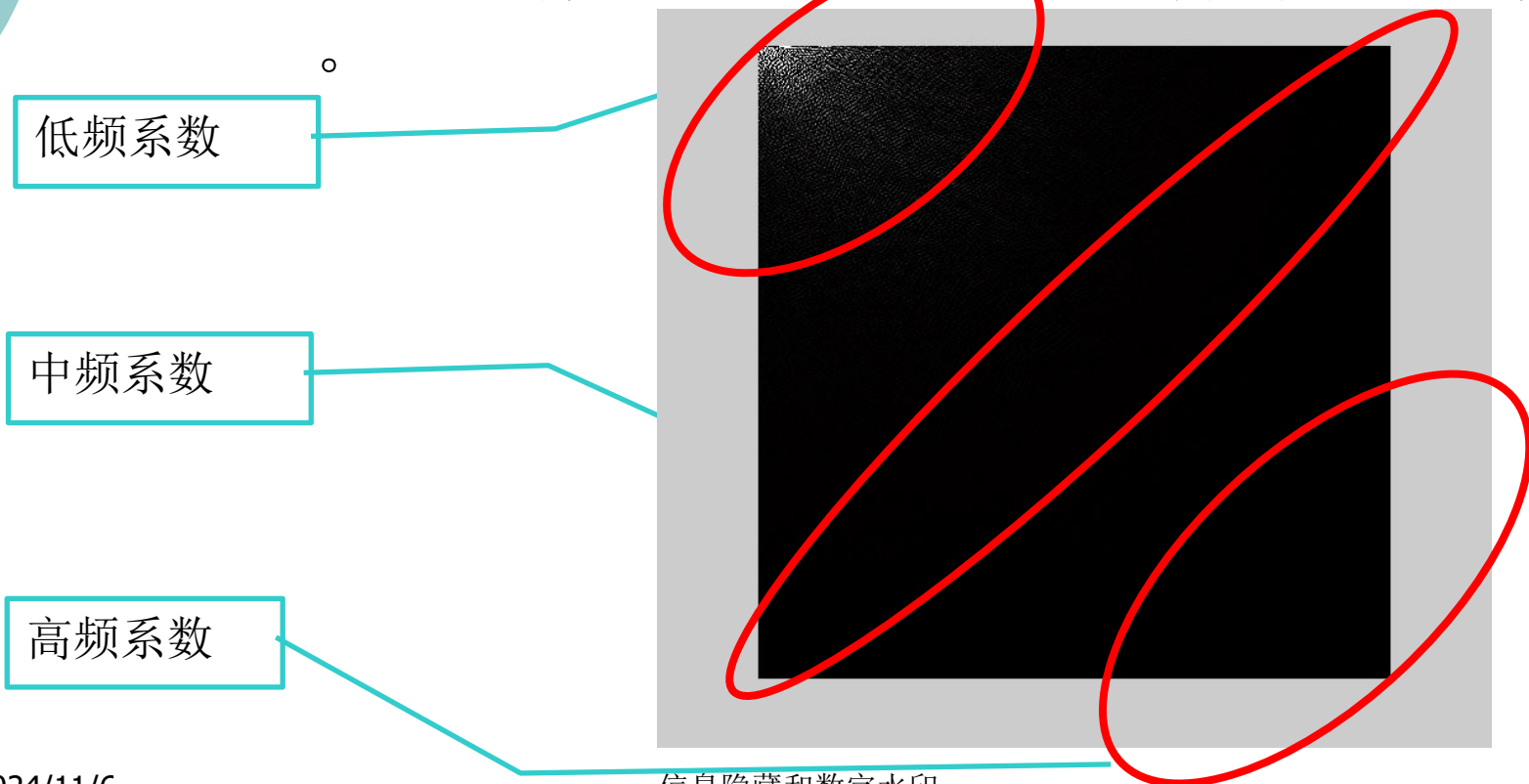
● 1. 准备图像信号



常用图像信号变换

○ 案例四：DCT系数特性研究

- 2. 对图像进行2维DCT变换，得到DCT系数



常用图像信号变换

○ 案例四：DCT系数特性研究

- 3. 分别用两组系数重构图像。第一组：原始系数（左图），第二组：部份高频系数置零（右图）



小结

- 数字音频是时间和样点值都离散的，计算机能处理的信号。
- 数字音频信息按一定的格式保存在磁盘文件中，通常包括采样率，编码方式，编码速率，单或双声道样点信息。
- 响度是人耳对声音强度的主观感受。人耳对不同频率的单音的敏感程度不同，物理强度相同的单音，可能因为频率不同，导致听起来响亮程度不同。

小结

- 人耳难以感知位于强信号附近的弱信号，这种声音心理学现象称为掩蔽。
- 强信号称之为掩蔽音，弱信号称之为被掩蔽音。掩蔽音和被掩蔽音同时存在所产生的掩蔽效应称为同时掩蔽或频域掩蔽，否则称之为异时掩蔽或时域掩蔽。

小结

- 数字图像是空间和样点值都离散的，计算机能处理的信号。
- 数字图像信息按一定的格式保存在磁盘文件中，通常包括分辨率，图像类型，调色板，样点值等信息。
- 常见图像包括二值、灰度、真彩色和调色板图像。
- 亮度是人眼对光强的主观感受。客观上相同的亮度，当平均亮度不同时，主观感觉的亮度也不相同。

小结

- 人由亮处走到暗处时的视觉适应过程，称为暗适应。人由暗处走到亮处时的视觉适应过程，称为亮适应。
- 人眼所看到的影像消失后，人眼仍能继续保留其影像约0.1-0.5秒左右。称为视觉惰性或暂留，电影和动画的制作利用了人眼的视觉惰性。

小结

- 质量评价包括主客观方法。MOS（Mean Opinion Score，平均意见分）为常用主观评价方法。
- SNR（Signal to Noise Ratio，信噪比）为音频质量常用客观评价方法。
- PSNR(Peak Signal to Noise Ratio，峰值信噪比)为图像质量常用客观评价方法。
- 客观评价与主观感受不一定相符。

小结

- 常用语音和图像信号处理操作包括DFT（Discrete Fourier Transform，离散傅里叶变换），DCT（Discrete Cosine Transform，离散余弦变换），DWT（Discrete Wavelet Transform，离散小波变换）。

小结

- DFT、DCT、DWT是一对可逆变换，正变换将信号从时域（音频样点）或空域（图像样点）转换为变换域（变换域系数），反变换则将变换域系数重构为音频或图像。
- 变换域系数的大小反映了信号不同频率分量能量的高低，变换域系数的位置与频率高低关联。通常修改高频系数所引起的失真不容易被察觉。

练习题

○ 填空题

○ 1、掩蔽效应分为_____和_____, 或_____和_____, 后者又分为_____和_____

A、同时掩蔽 B、时域掩蔽 C、频域掩蔽 D、超前掩蔽 E、滞后掩蔽 F、异时掩蔽

○ 2、人耳对声音强度的主观感受称为_____。

○ 3、已知图像分辨率为**1024*768**, 则图像每行有_____个像素, 每列有_____个像素。

练习题

○ 填空题

- 4、MOS是一种音频或图像质量主观评价方法，其英文全名为_____。
- 5、常见图像包括二值图像，_____, 真彩色图像，和_____。
- 6、人由亮处走到暗处时的视觉适应过程，称为_____。人由暗处走到亮处时的视觉适应过程，称为_____。

练习题

- 判断下列说法是否正确，正确填（T），错误填（F）。
 - 1、每秒种观察信号大小的次数，称为采样频率，或采样率。（ ）
 - 2、音频通常分为单声道和双声道两类，单声道音频能产生立体声效果。（ ）
 - 3、响度的单位为方，定义为1000Hz，10dB纯音的声强级。（ ）

练习题

- 判断下列说法是否正确，正确填（T），错误填（F）。
 - 4、MOS通常用3级评分标准来评价载体的质量。（ ）
 - 5、客观上相同的亮度，当平均亮度不同时，主观感觉的亮度仍然相同。（ ）
 - 6、修改高频系数导致的失真很容易被感知。（ ）

练习题

○ 计算题。

- 已知原始音频部分样点值如下：
- $x[i]: 10, 12, 14, 8, 6, 8$
- 隐藏信息后，该音频相应样点值变化为：
- $y[i]: 8, 13, 14, 9, 8, 6$
- 请计算这部份样点值序列的SNR。

练习题

○ 计算题。

- 已知原始图像部分像素值如下：
- $x[i,j]: 10, 12, 14,$
- $8, 6, 8$
- 隐藏信息后，该图像相应像素值变化为：
- $y[i,j]: 10, 12, 14,$
- $9, 8, 6$
- 请计算这部份像素构成的子图的PSNR。

参考文献

1. 人类视觉与色度学_百度文库
<http://wenku.baidu.com/view/d04fa51614791711cc791758.html>
2. http://blog.sina.com.cn/s/blog_4b72aa9d01008mvl.html
3. <http://tieba.baidu.com/f?kz=82344802>
4. 鈕心忻主编，信息隐藏与数字水印
5. matlab图像处理宝典，电子工业出版社



附录——习题解答

练习题解答

○ 填空题

- 1、掩蔽效应分为 A 和 F，或 C 和 B，后者又分为 D 和 E

A、同时掩蔽 B、时域掩蔽 C、频域掩蔽 D、超前掩蔽 E、滞后掩蔽 F、异时掩蔽

解：掩蔽信号和被掩蔽信号同时发生，称为**同时掩蔽**，也称**频域掩蔽**。掩蔽和被掩蔽信号非同时发生，称为**异时掩蔽**，也称**时域掩蔽**。若被掩蔽信号出现在掩蔽信号**之前**，则称为**超前掩蔽**，否则称为**滞后掩蔽**。

练习题解答

- 填空题
- 2、人耳对声音强度的主观感受称为响度。
- 3、已知图像分辨率为 $1024*768$ ，则图像每行有1024个像素，每列有768个像素。

练习题解答

○ 填空题

- 4、MOS是一种音频或图像质量主观评价方法，其英文全名为Mean Opion Score。
- 5、常见图像包括二值图像，灰度图像，真彩色图像，和调色板图像。
- 6、人由亮处走到暗处时的视觉适应过程，称为暗适应。人由暗处走到亮处时的视觉适应过程，称为亮适应。

练习题解答

- 判断下列说法是否正确，正确填（T），错误填（F）。
 - 1、每秒种观察信号大小的次数，称为采样频率，或采样率。（ T ）
 - 2、音频通常分为单声道和双声道两类，单声道音频能产生立体声效果。（ F ）
 - 解：双声道音频能产生立体声效果。

练习题解答

- 判断下列说法是否正确，正确填（T），错误填（F）。
 - 3、响度的单位为方，定义为1000Hz，10dB纯音的声强级。（ F ）
 - 解：频率为1000Hz，声强级为1dB的纯音的响亮程度被定义为一方。
 - 4、MOS通常用3级评分标准来评价载体的质量。（ F ）
 - 解：MOS通常用5级评分标准来评价载体的质量。

练习题解答

- 判断下列说法是否正确，正确填（T），错误填（F）。
 - 5、客观上相同的亮度，当平均亮度不同时，主观感觉的亮度仍然相同。（F）
 - 解：平均亮度影响我们对亮度的感受。
 - 6、修改高频系数导致的失真很容易被感知。（F）
 - 解：一般情况下，修改高频系数导致的失真不容易被感知。

练习题解答

○ 计算题。

- 已知原始音频部分样点值如下：
- $x[i]: 10, 12, 14, 8, 6, 8$
- 隐藏信息后，该音频相应样点值变化为：
- $y[i]: 8, 13, 14, 9, 8, 6$
- 请计算这部份样点值序列的SNR。

练习题解答

- 解：首先计算信号能量，根据公式可知：
- $p_s = \sum_{i=0}^5 x^2[i] = 10^2 + 12^2 + \dots + 8^2 = 604$
- 再计算噪声能量，原始音频和水印音频的差值即为噪声：
- $p_n = \sum_{i=0}^5 n^2[i] = \sum_{i=0}^5 (y[i] - x[i])^2 = (8 - 10)^2 + (13 - 12)^2 + \dots + (6 - 8)^2 = 14$
- 最后计算信号与噪声的比值
- $snr = p_s/p_n = 604/14 \approx 43.14$
- 换算为dB: $snr = 10\lg(43.14) \approx 16.35dB$

练习题解答

○ 计算题。

- 已知原始图像部分像素值如下：
- $x[i,j]: 10, 12, 14,$
- $8, 6, 8$
- 隐藏信息后，该图像相应像素值变化为：
- $y[i,j]: 10, 12, 14,$
- $9, 8, 6$
- 请计算这部份像素构成的子图的PSNR。

练习题解答

- 解：首先计算均方误差，**MSE**，由公式得：
- $MSE = \frac{1}{2 \times 3} \sum_{i=0}^1 \sum_{j=0}^2 (x[i, j] - y[i, j])^2 = \frac{1}{6} \times [(10 - 10)^2 + \dots + (6 - 8)^2] \approx 1.50$
- 再计算**PSNR**，以灰度图考虑，最大像素值为255，则：
- $PSNR = 10 \lg \left(\frac{x_{Max}^2[i, j]}{MSE} \right) = 10 \lg \left(\frac{255^2}{1.50} \right) \approx 43.37 dB$



附录——人类视觉其他特性

人类视觉特点

○ 视觉范围

- 视觉范围非常宽，但是人眼并不能同时感受这样宽的亮度范围。
- 当平均亮度比较适中时，能分辨的亮度的范围较大。
- 当平均亮度较低时，能分辨的亮度范围较小。

人类视觉特点

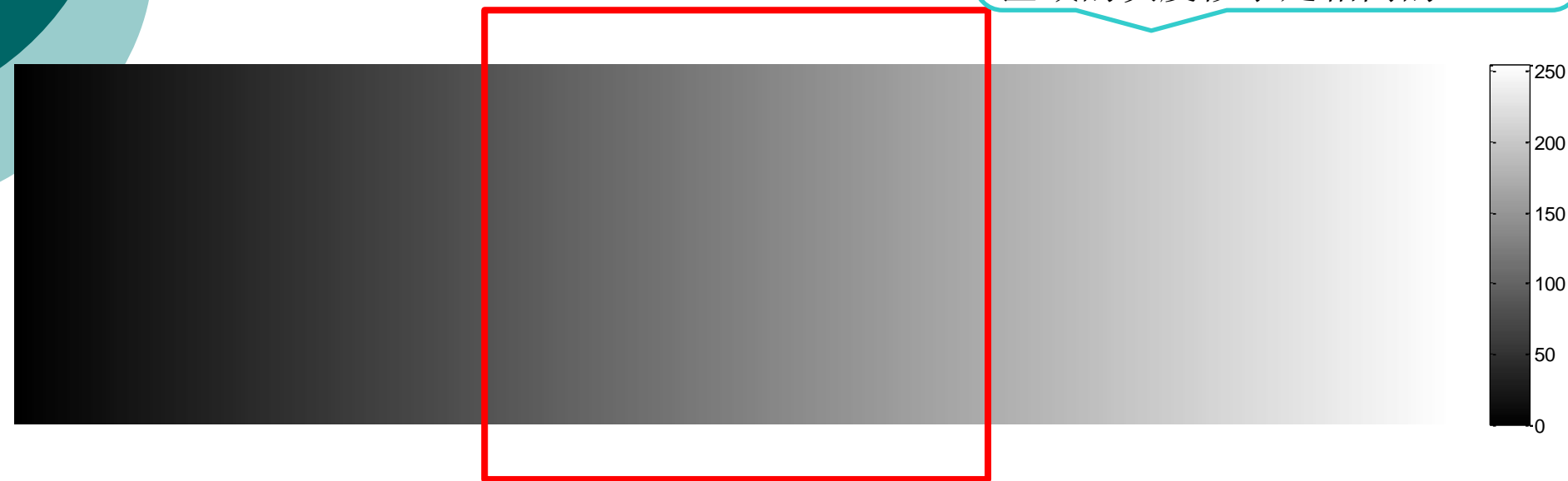
○ 对比灵敏度

● 可见度阈值

- 主观可辨别的亮度差所需的最小光强差。
- 即：当光强 I 增大时，在一定幅度内人眼感觉不出亮度变化；必须变化到一定值 $I+\Delta I$ 时，人眼才能感觉到亮度有变化。

人类视觉特点

红框宽度区域灰度跨越约为80个级别，但在我们看来，整个区域的灰度似乎是相同的。



人类视觉特点

○ 对比灵敏度

● 对比灵敏度

- 可见度阈值与光强比值，即 $\Delta I/I$ 。

● 误差低于对比灵敏度，不会被人眼察觉。

- 图像编码量化器利用这一视觉特性：图像的边缘可以容忍较大的量化误差，因而可减少量化级，从而降低数字码率。

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 考虑均匀量化，幅值是 $[0, 127]$ 上的实数。
- 若量化单位为2，量化系数采用四舍五入法
 - 则量化系数需要多少比特表示？
 - 解：幅值动态范围为 $[0, 127]$ ，区间宽度为128，若量化单位为2，则可划分为64个量化级别，所以最多需要6比特即可表示量化系数。

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 考虑均匀量化，幅值是 $[0, 127]$ 上的实数。
- 若量化单位为2，量化系数采用四舍五入法
 - 若幅值为7，则量化系数为？
 - 解： $\text{round}(7/2)=\text{round}(3.5)=4$ 。（注：
`round`表示四舍五入函数）

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 考虑均匀量化，幅值是 $[0, 127]$ 上的实数。
- 若量化单位为2，量化系数采用四舍五入法
 - 由量化系数重构的幅值为？量化误差为？
 - 解：重构的幅值 = 量化系数 \times 量化单位 = $4 \times 2 = 8$
； 则量化误差为 = 重构幅值 - 原始幅值 = $8 - 7 = 1$

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 若量化单位为4，幅值为6，则上述问题解答有何变化？
 - 解：若量化单位为4，则
 - 1) 幅值动态范围为 $[0, 127]$ ，区间宽度为128，则可划分为32个量化级别，所以最多需要5比特即可表示量化系数。

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 若量化单位为4，幅值为6，则上述问题解答有何变化？
 - 解：若量化单位为4，则
 - 2) 量化系数为： $\text{round}(6/4)=\text{round}(1.5)=2$
。（注：round表示四舍五入函数）

人类视觉特点

——对比灵敏度与码速率

○ 编码速率和量化误差

- 若量化单位为4，幅值为6，则上述问题解答有何变化？

- 解：若量化单位为4，则

- 3) 重构的幅值 = 量化系数 \times 量化单位 = $2 \times 4 = 8$
； 则量化误差为 = 重构幅值 - 原始幅值 = $8 - 6 = 2$

人类视觉特点

——对比灵敏度与码速率

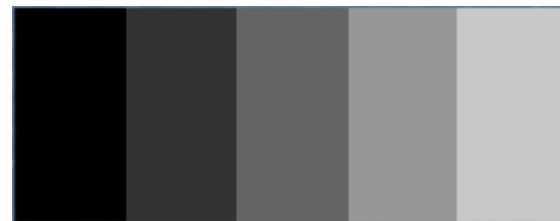
- 从中可以得出什么结论？
 - 量化粒度越细，即量化单位越小，量化误差越小
 - 然而量化单位越小，同时意味着量化级越多，
 - 即表示量化系数所需的比特数增加，直接导致码速率增加。
 - 利用视觉特性，在人眼不敏感区域，量化粒度粗，既可以降低总体码速率，又不影响感官效果。

人类视觉特点

○ 马赫效应

- 当亮度发生跃变时，会有一种边缘增强的感觉。
- 视觉上会感到亮侧更亮，暗侧更暗。
- 边缘效应：图示五个灰度级，带内亮度相同。

（图片源自参考文献1）



人类视觉特点

○ 马赫效应

- 马赫效应会导致局部阈值效应。
- 即在边缘的亮侧，靠近边缘像素的误差感知阈值比远离边缘阈值高 $3\sim 4$ 倍，因此这些像素的编码误差可以大一些。

人类视觉特点

○ 分辨力

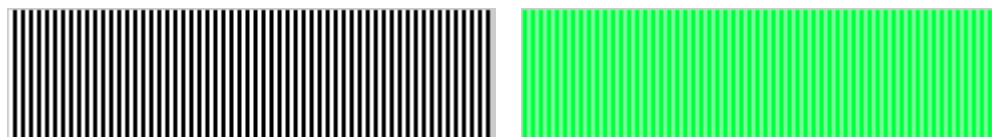
- 人眼在一定距离上能区分开相邻两点的能力。
- 环境照度影响：照度太低和太高，分辨力降低。
- 物体的运动速影响：速度大，则分辨力下降。

人类视觉特点

○ 分辨力

- 人眼对彩色的分辨力要比对黑白的分辨力低。
 - 当把对黑白细节的分辨能力定义为100%时，人眼对其他彩色的分辨力：
 - 黑白：100，黑绿：94，黑红：90，黑蓝：26，红绿：40，红蓝：23，蓝绿：19（数据源自参考文献1）

人类视觉特点





附录——图像质量评价指标

参考文献

图像视觉质量评价基础知识

- <https://www.cnblogs.com/ranjiewen/p/6390846.html>

- 标准图像库

- <http://live.ece.utexas.edu/research/Quality/>

- IQA指标分析

- <http://sse.tongji.edu.cn/linzhang/IQA/IQA.htm>

- 包含常用IQA实现，以及IQA在标准图像库上的性能分析。

- 视频质量专家组（Video Quality Experts Group , VQEG）

- <https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>