

信息隐藏技术基础

——感知影响的估计以及感知模型

主要内容

感知影响的估计

感知模型的一般形式

感知影响的估计——保真度和质量

如何度量图像的感知可见性？

两类特性：

✓ *Fidelity*（保真度）

✓ *Quality*（质量）

感知影响的估计——保真度和质量

Fidelity is a measure of the similarity **between signals** before and after processing.



Original Work



**Add Gaussian Noise
Amount: 1%**

High-fidelity reproduction



**Add Gaussian Noise
Amount: 30%**

Low-fidelity reproduction

感知影响的估计——保真度和质量

Quality is an **absolute** measure of appeal.

A high-quality image or high-quality audio clip simply looks or sounds good.

High-quality



Original Work

Medium-quality



Add Gaussian Noise
Amount: 1%

Low-quality



Add Gaussian Noise
Amount: 30%

感知影响的估计——保真度和质量

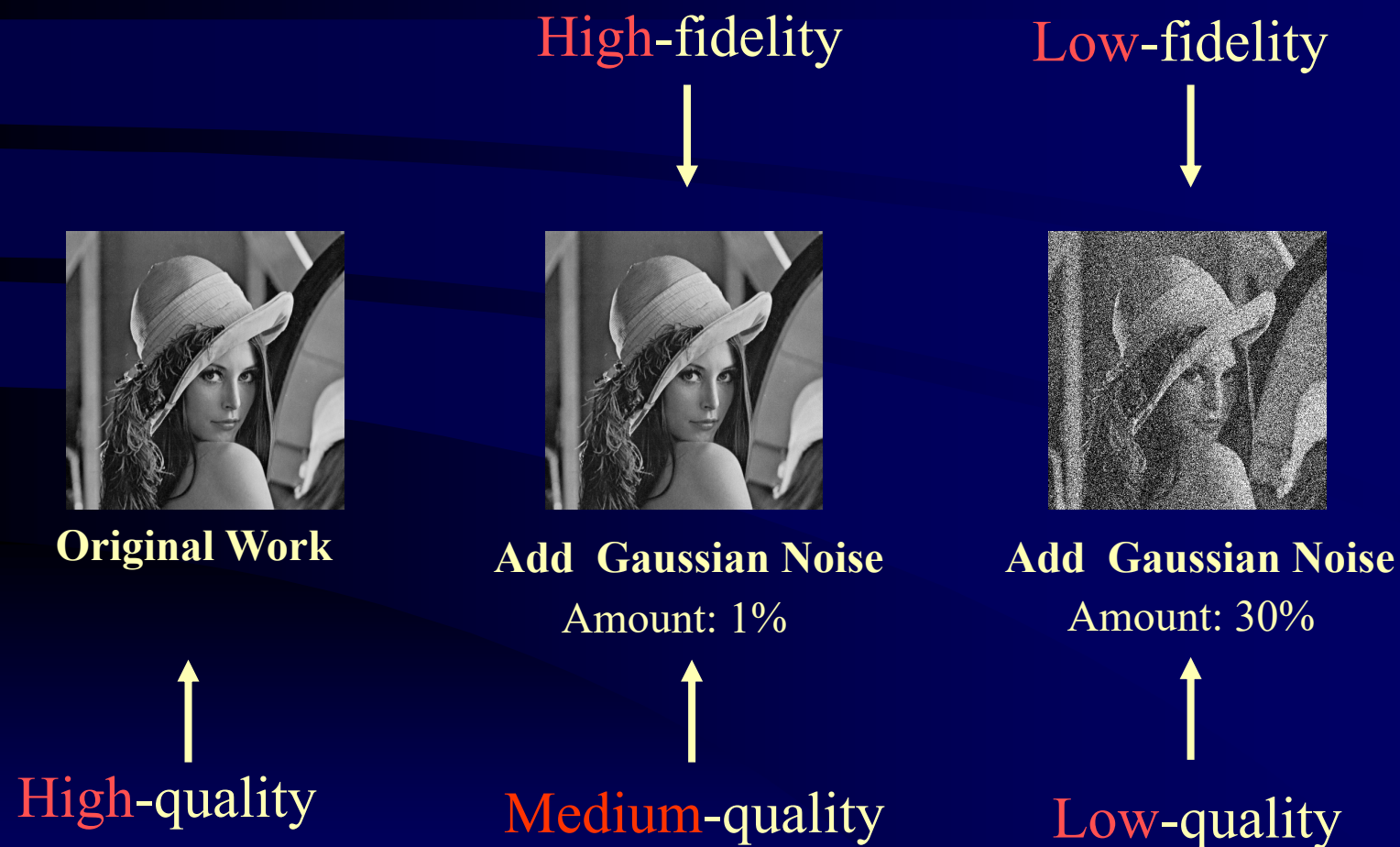
High fidelity \neq High quality



The **difference** between fidelity and quality

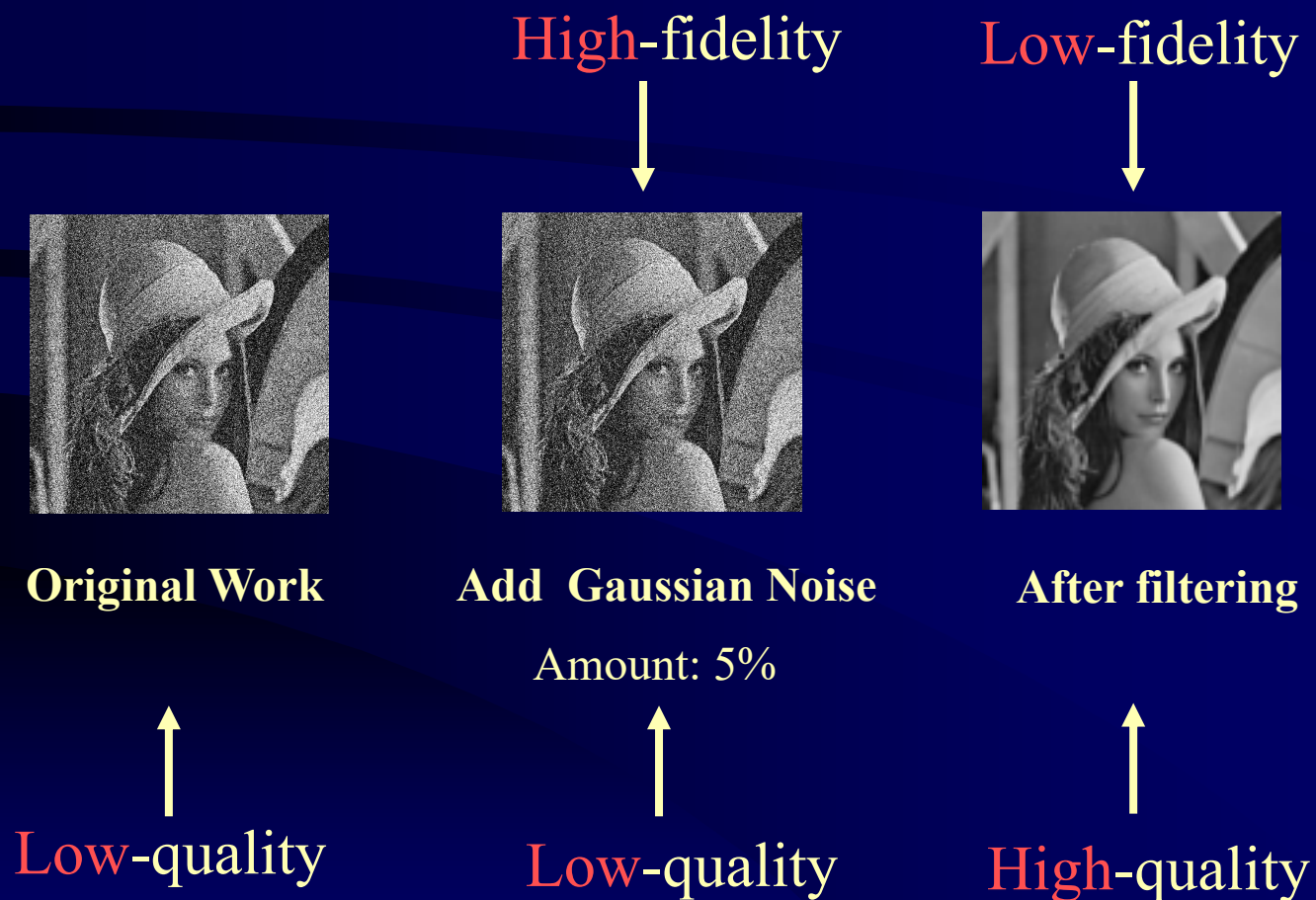
感知影响的估计——保真度和质量

Example 1



感知影响的估计——保真度和质量

Example 2



感知影响的估计——保真度和质量

在一些场合，保真度是评价含隐藏信息图像好坏的重要指标，目的是尽量保持原始图像与处理后的图像的一致性！

对于一些出土的珍贵图像，数字化后其图像质量由于年代的久远，不是很好，但用于其版权保护的数字水印的嵌入，必须考虑其保真度！



↑
原始图像

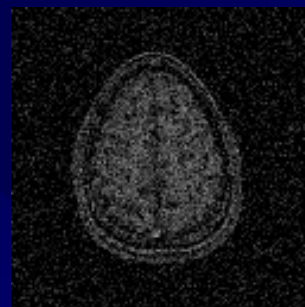
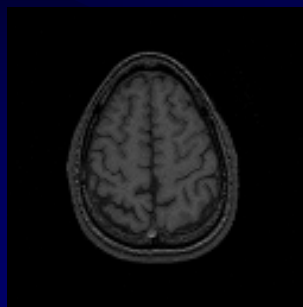


↑
含水印信息图像



↑
水印信息

医学图像的保护与认证也以保真度为前提的！



感知影响的估计——保真度和质量

在另一些场合下，图像的质量是重要的评价指标，比如，一个现代艺术品图像的保真度等。



感知影响的估计

——感知评估问题

感知影响的估计——感知评估问题

基于人感知的评估技术

尽管图像的感知影响估计一直在讨论，但是基于人的感知的图像的保真度与图像质量的评价指标的研究还远远不够。通常利用的是不同人群的民意测验的方式进行，但是由于数量总是有限的，很难做到准确的自动评估（ **automated evaluations** ）。

感知影响的估计——感知评估问题

两个基于统计方法的有用的主观评估准则——

❖ The *two alternative, forced choice* (2AFC)

——二选一方法

❖ Rating the quality of a Work (ITU-R Rec. 500)

——质量分级方法

感知影响的估计——感知评估问题

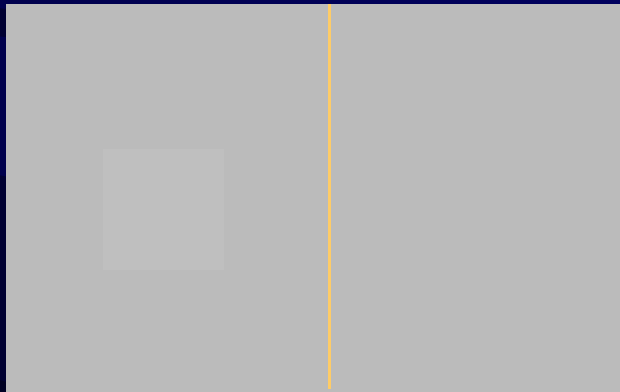
- ◆一些研究表明，在人类的视听感的判断中，人与人之间的区别是极大的。
- ◆这种感觉在个体也是不断变化的
- ◆但是，如果主观和客观的测试量极大的话，还是能得到一些普遍的结论的。尤其是对于特定的群体而言，可以得到较一致的结论性的内容的 (比如18到25岁年龄段的实验数据)。
- ◆这种大量的主观感知的评估为客观（自动）感知信息的评估提供了指导性的依据

感知影响的估计——感知评估问题（2AFC）

❖ *The two alternative, forced choice (2AFC)*

感知影响的估计——感知评估问题（2AFC）

下面是一个 *two alternative, forced choice (2AFC)* 测验的例子。



Is the square on left or right?

被测者必须做出选择，哪怕是猜

感知影响的估计——感知评估问题（2AFC）



图像质量的测验



Which one has higher quality ?

感知影响的估计——感知评估问题（2AFC）



Original



A



B

2AFC 也可以作为保真度的测验

被测者被问及A或B哪一个图像更像原始图像？

感知影响的估计——感知评估问题（分级准则）

❖ 图像质量分级测验方法（ **rate the quality of a Work** , **ITU-R Rec. 500** ）。

感知影响的估计——感知评估问题（分级准则）

1. 在ITU-R Rec. 500测验中的分级表

Five-Grade Scale			
Quality（质量）		Impairment（损伤度）	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

感知影响的估计——感知评估问题（分级准则）

2. 准备图像数据库

◆例如：测试图像选择 N 张经典的基准图像

这些图像包含各种各样的类型，包括动物，汽车，飞机，人像，复杂的自然环境以及没有特别的感兴趣主体对象的图像等。



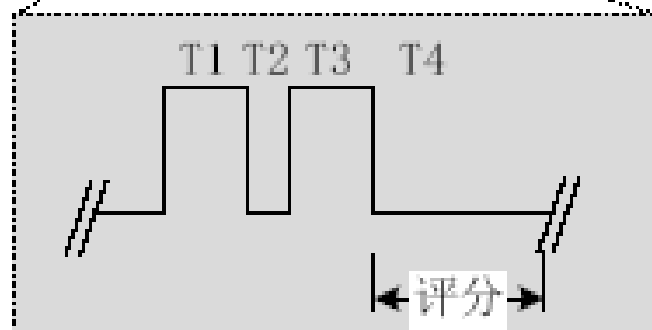
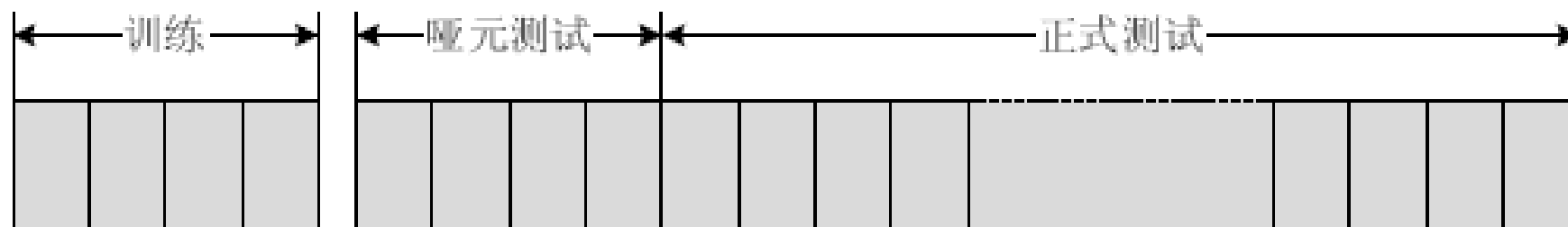
感知影响的估计——感知评估问题（分级准则）

- ◆每个图像都是 512×512 像素的灰度图
- ◆图像失真：这 N 个参考图（原图）被6种有代表性的攻击手段攻击，一共产生 $N \times 6$ 个受攻击的目标图
- ◆失真类型：JPEG压缩，旋转，直方图均衡化，两种高斯噪声（分别由均值或方差控制）以及马赛克
- ◆攻击强度：要覆盖很宽的范围，比如从95%的JPEG压缩到方差为0.95的高斯噪声。

感知影响的估计——感知评估问题（分级准则）

3. 测试会话

测试会话由一个程序自动控制，包括准备、训练、哑元测试和正式测试几个部分，大约持续30—40分钟。



展示的阶段划分:

T1 = 10s 参考图像

T2 = 3s 灰屏

T3 = 10s 目标图像

T4 = 5s 质量/失真评分表

A Presentation

感知影响的估计——感知评估问题（分级准则）

◆准备

在正式测试之前，评价员将被告知测试的目的和方法，并被要求依照图像检索的场景来熟悉和理解前表规定的评分等级。

◆训练

- 示例程序被用来让评价员熟悉测试程序的流程
- 在示例程序中，5个不同的参考图和5个不同强度攻击的目标图被按照正式测试的逻辑呈现，并要求评价员按照正式测试来给出评分
- 评价员在示例程序中的评分不计入最终的测试结果，但可以用来确认他对试验的目的和方法有了正确的认识
- 在示例程序中出现的参考图不会在正式测试中出现。

感知影响的估计——感知评估问题（分级准则）

◆哑元测试

- 在训练结束之后，评价员被告知进入正式测试
- 但事实上，最初几个测试结果作为哑元测试，其结果并不被程序记录，只是为了让评价员适应测试的环境。

◆正式测试

- 由多个循环的展示（Presentation）构成，每个展示随机的从测试数据库中选择一个参考图，并相应的随机选择由某个攻击方式和某个攻击强度产生目标图，得到一个评分记录。
- 在随机选择中，相同的参考图和相同的攻击方式都不会连续出现两次
- 一个展示包括四个时间固定的阶段
- 评分阶段的时延为5秒，在这5秒之内，评价员必须给出目标图质量/失真的打分
- 如果由于评价员的犹豫错过了这个时间，测试程序将自动进行下一次循环，并放弃这一次打分的记录。

感知影响的估计——感知评估问题（分级准则）

◆评分：

➤对于同一测试条件下的 N 个测试样本，其均值、标准差为：

$$U = \frac{1}{N} \sum_{i=1}^N r_i$$
$$S = \sqrt{\sum_{i=1}^N \frac{(U - r_i)^2}{N - 1}}$$

其中， r_i 为一个主观评价得分， N 是样本个数。

由此， $\varepsilon = 1.96 \frac{S}{\sqrt{N}}$ ，则可以将落在95%置信区间 $[U - \varepsilon, U + \varepsilon]$ 之外的评分视为出界评分。对于出界的评分，将相应的测试混在其它正常的测试中执行补充试验，直到获得合格的测试数据。

主要内容

感知影响的估计

感知模型

人类视觉系统

人类视觉系统（**Human Visual System, HVS**）的响应随着输入信号空间频率、亮度和色彩的变化而变化，它表明载体图像各区域或各频带的感知程度并不是等同的。感知程度上的不同可以测量出来，同时也可构造出一些模型对这些差异进行解释。一般来说，一个感知模型涉及三个基本概念：灵敏度、掩蔽效应和综合（**Pooling**）。

人类视觉系统---灵敏度

灵敏度是指耳朵或眼睛对于直接激励的反应程度。在一些灵敏度测定实验中，为观察者提供单独的激励并测试他们对于这些激励的感知程度。

例如，通常需要对最小的声音强度进行测量，只有达到最小声音强度才能听见某一特定频率的声音，且通常要在一个频率范围内重复进行测量。

视觉测量的最主要特性为：*频率灵敏度和亮度灵敏度*

人类视觉系统---灵敏度

频率灵敏度

人类听视觉系统对于某一输入信号的响应依赖于信号的频率。对于听觉来说，不同频率的信号会被感知成不同的音调。对于视觉来说，存在三种形式的频率响应。

1. 空间频率
2. 谱频率
3. 时间频率

人类视觉系统---灵敏度

空间频率被感知成图案或纹理。可以通过对比度灵敏度函数表示。

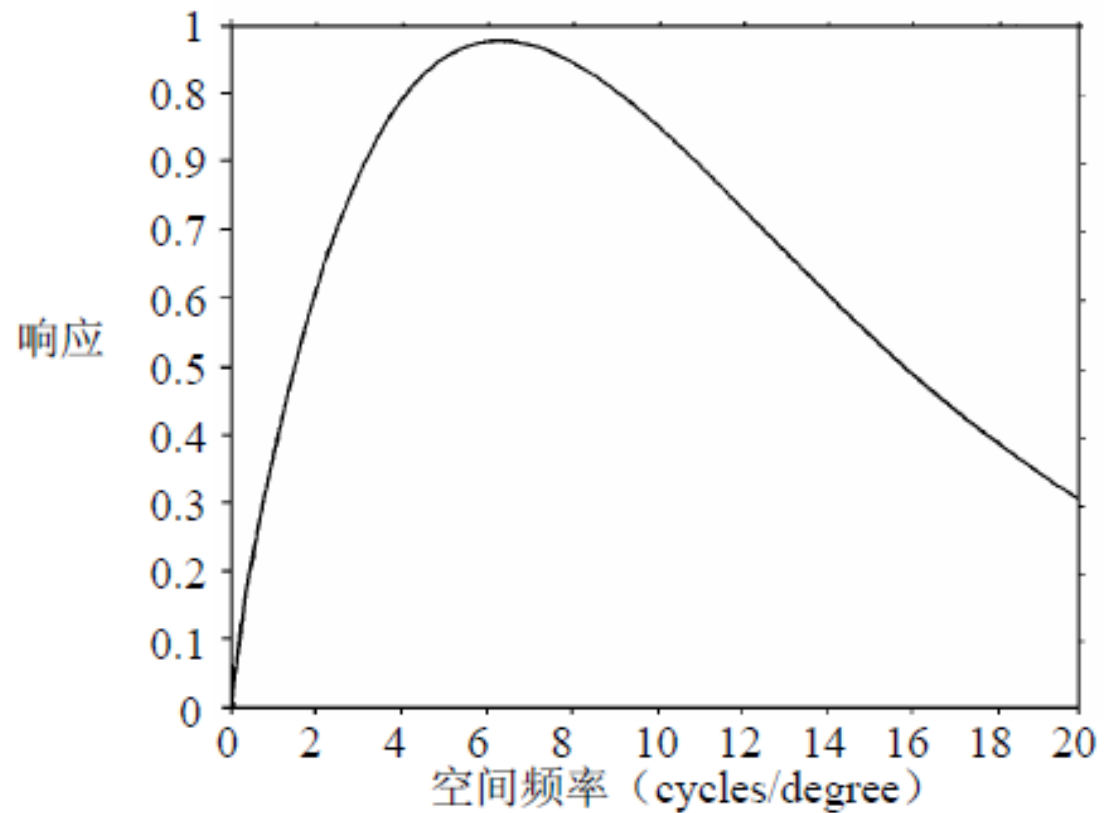
二维空间频率的描述可用它们的幅度和方位来表示。人眼的敏感程度并不仅仅依赖于图像的频率，而且也依赖于它们的方位。

人眼对于图像中水平和垂直的线和边缘最敏感，而对于成45度的线和边缘最不敏感。



人类视觉系统---灵敏度

对比度灵敏度函数



这幅图清楚地表明人们对于中频范围内的亮度变化最敏感，
而且在较低和较高的频率处人类的灵敏度会下降。

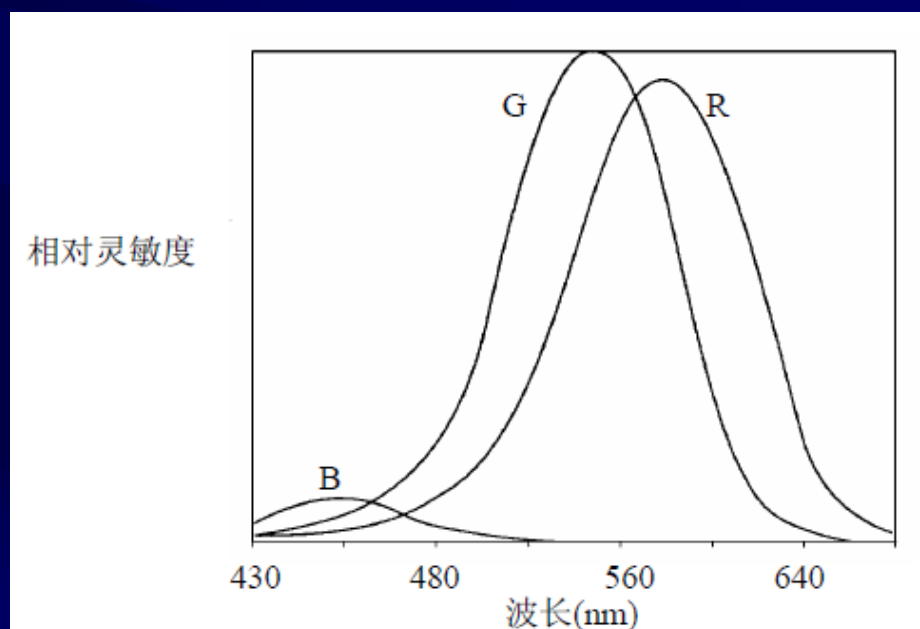
人类视觉系统---灵敏度

谱频率被感知为色彩。人眼对于彩色的感知，其最低一级是由三个单独的颜色通道组成的。

蓝色通道明显要比其余两个通道低。

一些彩色水印处理系统将大部分的水印信号嵌入到彩色图像的蓝色通道中。

三个颜色通道的相对灵敏度对比



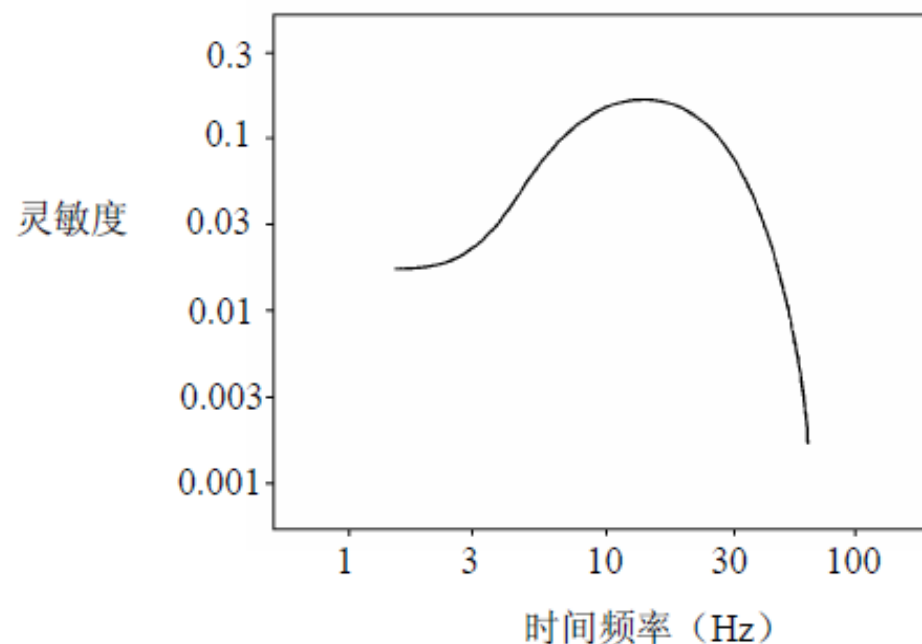
人类视觉系统---灵敏度

时间频率被感知成运动或闪烁。图中给出了一个实验的结果，这个实验用来测量人眼对于不同时间频率的反应。

实验结果显示当频率超过30Hz 时人眼的灵敏度下降非常快。

电视和电影的帧速率都不超过每秒60 帧的原因。

人类视觉系统对时间频率的灵敏度



人类视觉系统---灵敏度

相对于强度较小的声音信号，人耳对于强度较大的声音信号中的变化更为敏感。

然而对于人眼来说，情况恰恰相反，人眼对于亮度越高的信号越不敏感。亮度灵敏度是非线性的，并且已经通过对数关系、立方根关系以及一些更为复杂的模型对此进行了建模。

人类视觉系统---掩蔽效应

背景环境会影响人类的感知效果。单独的一个纹理特征很容易被发现，但是当这个纹理被加到一幅纹理复杂的图像中时它就有可能很难被发现。这也就是说，一个信号的存在能够隐藏或掩蔽另一个信号的存在。

掩蔽效应是一种测度，它用来测量观察者在“掩蔽”信号存在的情况下对某一激励的反应。

例如，在一幅图像加入均匀噪声后，虽然噪声是均匀的，但在这幅图像中噪声的可见性却非常不均匀，它明显地依赖于局部图像结构。在相对平滑的区域，噪声明显可见。相比之下，在纹理复杂部分，噪声几乎完全不可见。

人类视觉系统---掩蔽效应

掩蔽现象普遍存在。对于视觉来说，两个主要的例子就是频率掩蔽和亮度掩蔽。在频率掩蔽中，某一频率成分的存在能够掩蔽人类对于作品中另一频率成分的感知。在亮度掩蔽中，局部的亮度特征能够掩盖对比度的变化。

对于听觉来说，同样存在类似的频率和响度掩蔽条件。除此以外还存在着时域掩蔽。在这种掩蔽中，人们对于某一个声音信号的感知也许会被前一个或者后一个声音信号所掩蔽。

人类视觉系统---掩蔽效应



人类视觉系统---综合

灵敏度模型和掩蔽模型可用于对特定条件下（例如，单一频率）某一变化的感知情况进行估计。

如果是多个频率而不是单一频率发生变化，这时就需要将每个频率下的灵敏度特性和掩蔽特性的信息结合起来。

在感知距离模型中，将对各个失真的感知特性结合起来就得到有关作品总体变化的一个估计。这过程称为综合（**Pooling**）。

人类视觉系统---综合

通常采用下面这种形式的公式进行综合：

$$D(\mathbf{x}, \mathbf{x}^w) = \left(\sum_i |d_i|^p \right)^{\frac{1}{p}}$$

其中， d_i 是对某种可能性的一个估计。这种可能性是指观察者能否观察到 \mathbf{x} 和 \mathbf{x}_w 之间针对某一单个参数所存在的差异。参数可以是时域采样、空间像素或是傅立叶变换系数。

对于音频信号来说，一般 $p = 1$ ；

对于图像来说， $p = 4$ 更为常见。

人类视觉系统---*Watson* 感知模型

利用灵敏度、掩蔽以及综合的思想，现在来描述
Watson 提出的一个用来测量视觉逼真度的模型。

Watson 所提出的模型采用了分块DCT变换。

图像用 x 来表示， x_{ijk} 来表示第 k 块中的第 i, j 个像素。

用 X_{ijk} 来表示第 k 个像素块经过DCT变换后所得到的
DCT系数。

优点：用该模型评价含噪声图像的感知效果远远优于
PSNR。

人类视觉系统---Watson 感知模型

灵敏度

灵敏度特性模型定义了一个频率灵敏度表 t 。表中的每一项 t_{ij} 大约等于块中相应的DCT 系数在没有掩蔽噪声情况下无法被分辨出来的最小值。灵敏度表与许多参数有关，包括图像的分辨率、观察者到图像的距离等

DCT 频率灵敏度表

1.40	1.01	1.16	1.66	2.40	3.43	4.79	6.56
1.01	1.45	1.32	1.52	2.00	2.71	3.67	4.93
1.16	1.32	2.24	2.59	2.98	3.64	4.60	5.88
1.66	1.52	2.59	3.77	4.55	5.30	26.28	7.60
2.40	2.00	2.98	4.55	6.15	7.46	8.71	10.17
3.43	2.71	3.64	5.30	7.46	9.62	11.58	13.51
4.79	3.67	4.60	6.28	8.71	11.58	14.50	17.29
6.56	4.93	5.88	7.60	10.17	13.51	17.29	21.15

人类视觉系统---Watson 感知模型

亮度掩蔽特性

亮度自适应性是指如果 8×8 像素块的平均亮度更亮，那么可以对一个DCT 系数改变较大的量而不被人注意。Watson 针对每一小块，根据其直流分量大小来调整灵敏度表 t 。亮度掩蔽阈值 t_{ijk} 由下面的式子获得：

$$t_{ijk} = t_{ij} (X_{00k} / \bar{X})^{a_T}$$

其中 a_T 是一个常数，建议取值为0.649， X_{00k} 代表原始图像第 k 个像素块的直流系数， \bar{X} 为图像中所有直流系数的平均值。

表明图像中亮度越高的区域可承受较大的变化而不被人注意。

人类视觉系统---*Watson* 感知模型

对比度掩蔽特性

对比度掩蔽特性（即某一频率成分的能量使得另一频率成分的变化
的可见性下降）则会产生一个掩蔽门限值 s_{ijk}
其定义如下：

$$s_{ijk} = \max \{ t_{ijk}, |X_{ijk}|^{\beta_{ij}} \cdot (t_{ijk})^{1-\beta_{ij}} \}$$

其中 β_{ij} 是0 到1 之间的一个常数。对于所有的*i*和*j*，**Watson** 取 $\beta_{ij} = 0.7$ 。
。

人类视觉系统---Watson 感知模型

综合

为了比较原始图像 x 和对应的失真图像 x^w ，首先计算相应DCT 系数之间的差值 $e_{ijk} = X_{ijk}^w - X_{ijk}$ 。然后计算感知距离 d_{ijk} ：

$$d_{ijk} = \frac{e_{ijk}}{S_{ijk}}$$

综合感知距离：

$$D_{wat}(x, x^w) = \left(\sum_{i,j,k} |d_{ijk}|^p \right)^{\frac{1}{p}}$$

人类视觉系统---*Watson* 感知模型



感知距离: $D=124$

人类视觉系统---*Watson* 感知模型



感知距离: $D=98$

Thank you!

-
1. 质量和保真度的相同点和不同点是什么？
 2. 主观测试的具体步骤是什么？
 3. 视觉测量的最主要特性是什么？
 4. 人们对红绿蓝三种颜色中的哪种颜色最敏感？
 5. 什么是掩避效应？
 6. 建立Watson模型的目的什么？