


Quality of Experience



多媒体通信网络中的用户体验质量管理

 2019年11月10日 

《多媒体通信》课程专题

电子工程与信息科学系



引言：有限资源 \leftrightarrow 无限增长的需求

截至2018年6月，中国国际出口带宽为8,826,302Mbps

中国国际出口带宽及增长率

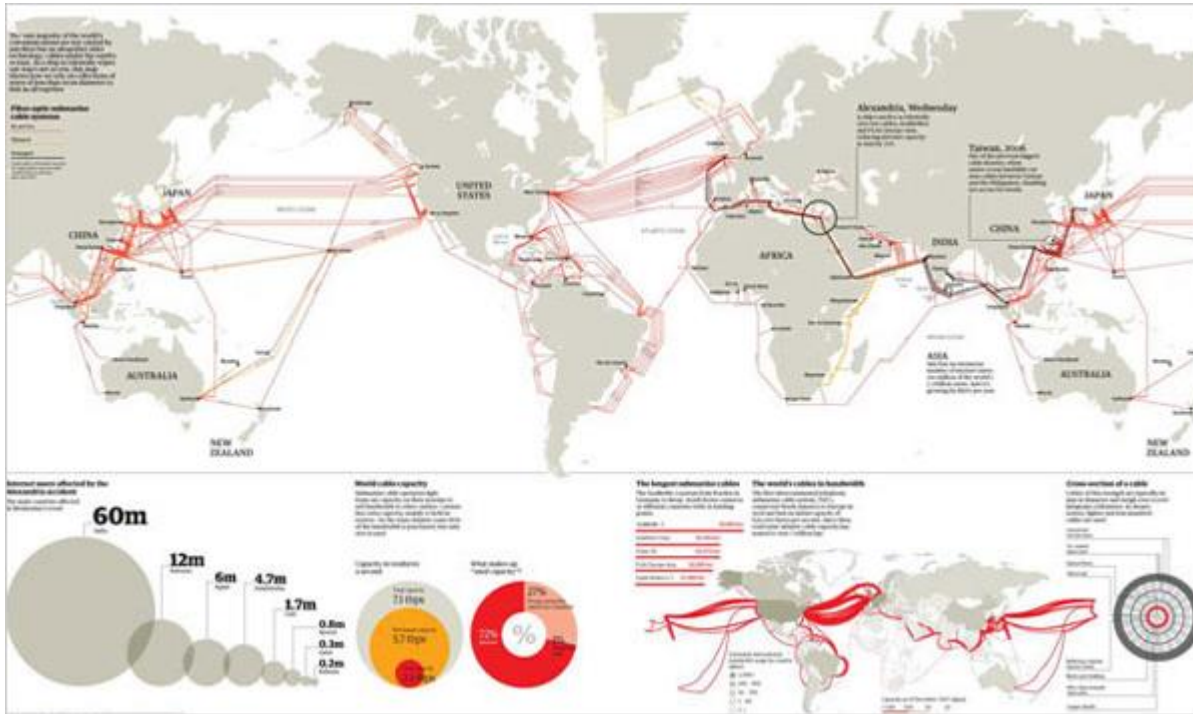
单位：Mbps





引言：有限资源 \leftrightarrow 无限增长的需求

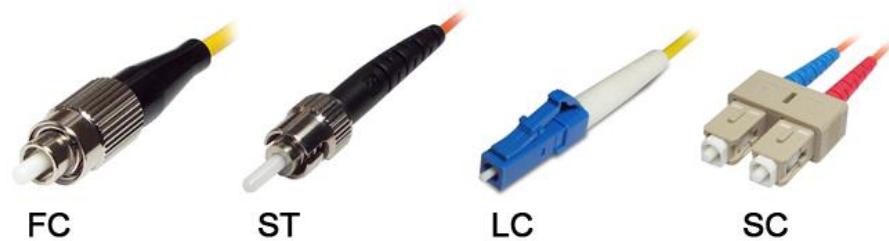
越洋光缆、光纤到户现状



截至2014年，海底通讯光缆数量已达到285条，其中22条不再使用，被称之为“黑光缆”。



截止2016年6月，中国各地光纤网络覆盖家庭数已超过50%



FC

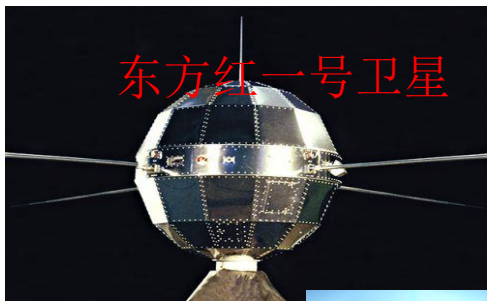
ST

LC

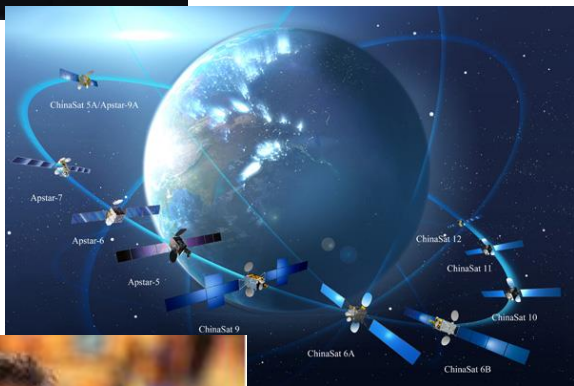
SC



引言：有限资源 \leftrightarrow 无限增长的需求 卫星通信/广播现状



东方红一号卫星：1970年4月24日，中国第一颗人造地球卫星在酒泉卫星发射中心成功发射，使中国成为继苏、美、法、日之后世界上第五个独立研制并发射人造地球卫星的国家。



截至2015年8月31日，地球轨道上共有1305颗正常工作的卫星，其中美国549颗，中国142颗，俄罗斯131颗，德国24颗。美国卫星中，152颗是军用卫星，250颗商用卫星。

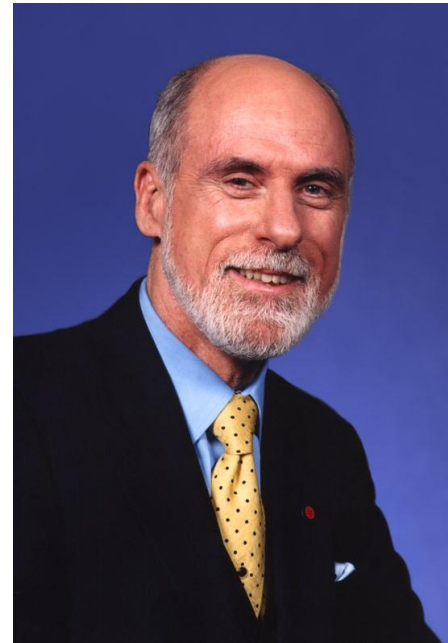
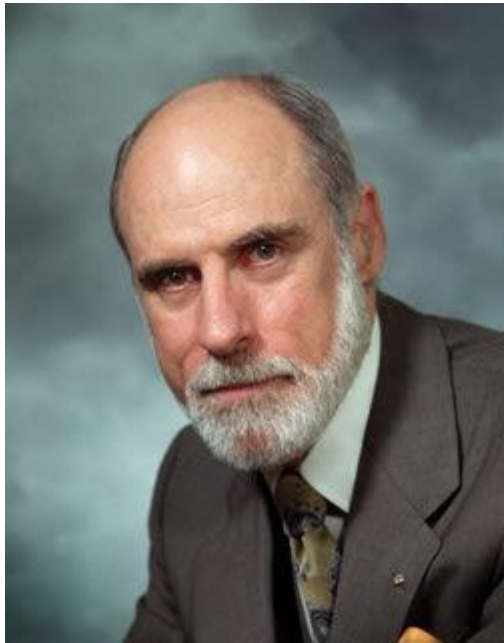
2016年8月6日，我国卫星移动通信系统首发星天通一号01星发射升空

2016年8月16日，“墨子号”发射



引言：有限资源 \leftrightarrow 无限增长的需求 网络还能发展吗？

◆ **Vint Cerf: We Still Have 80 Per Cent of the World to Connect. 2009**

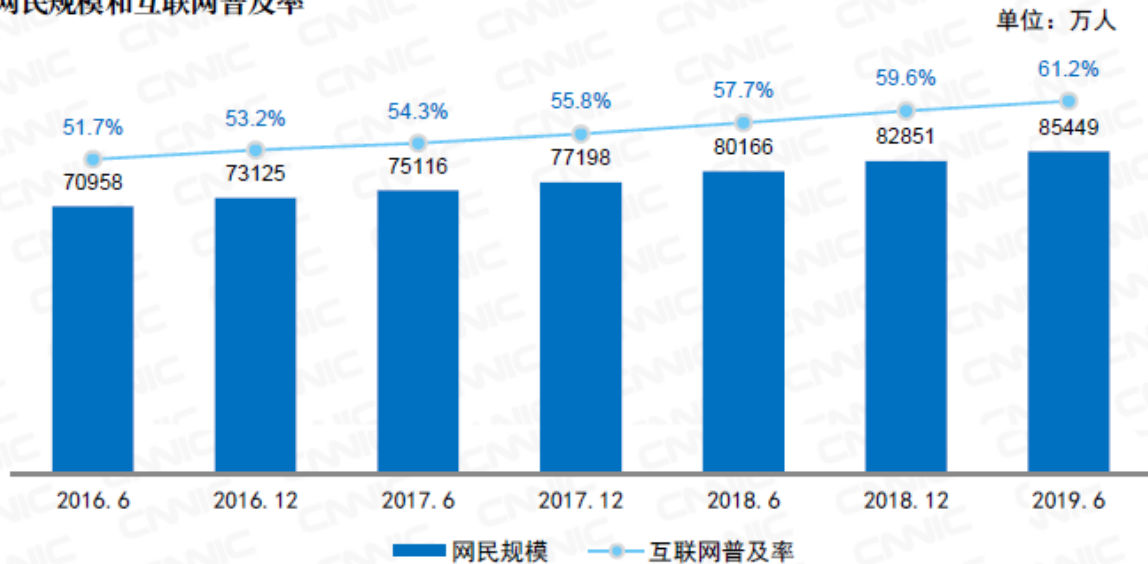




引言：有限资源 \leftrightarrow 无限增长的需求

移动互联网的现状

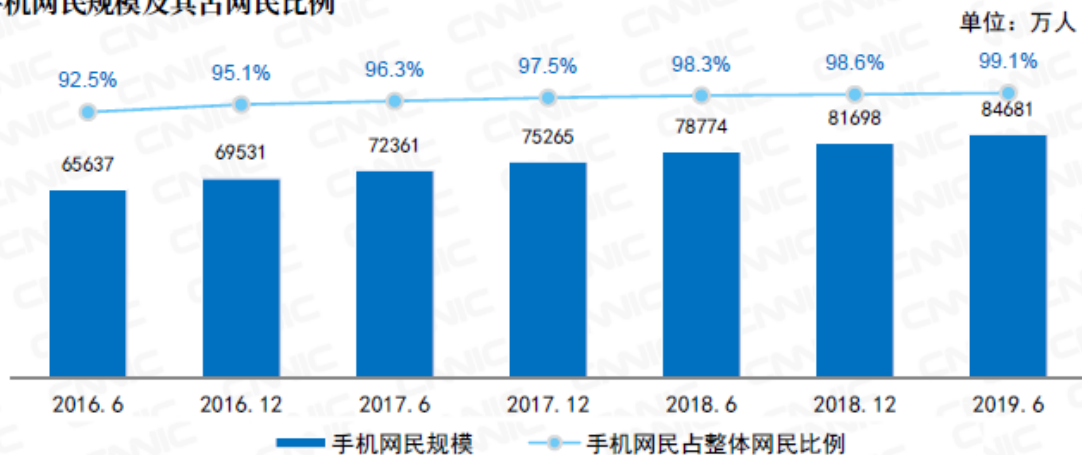
网民规模和互联网普及率



截至 2019年6月，我国网民规模达**8.54亿**

来源：CNIC 中国互联网络发展状况统计调查

手机网民规模及其占网民比例



截至 2019年6月，我国手机网民规模达**8.47亿**

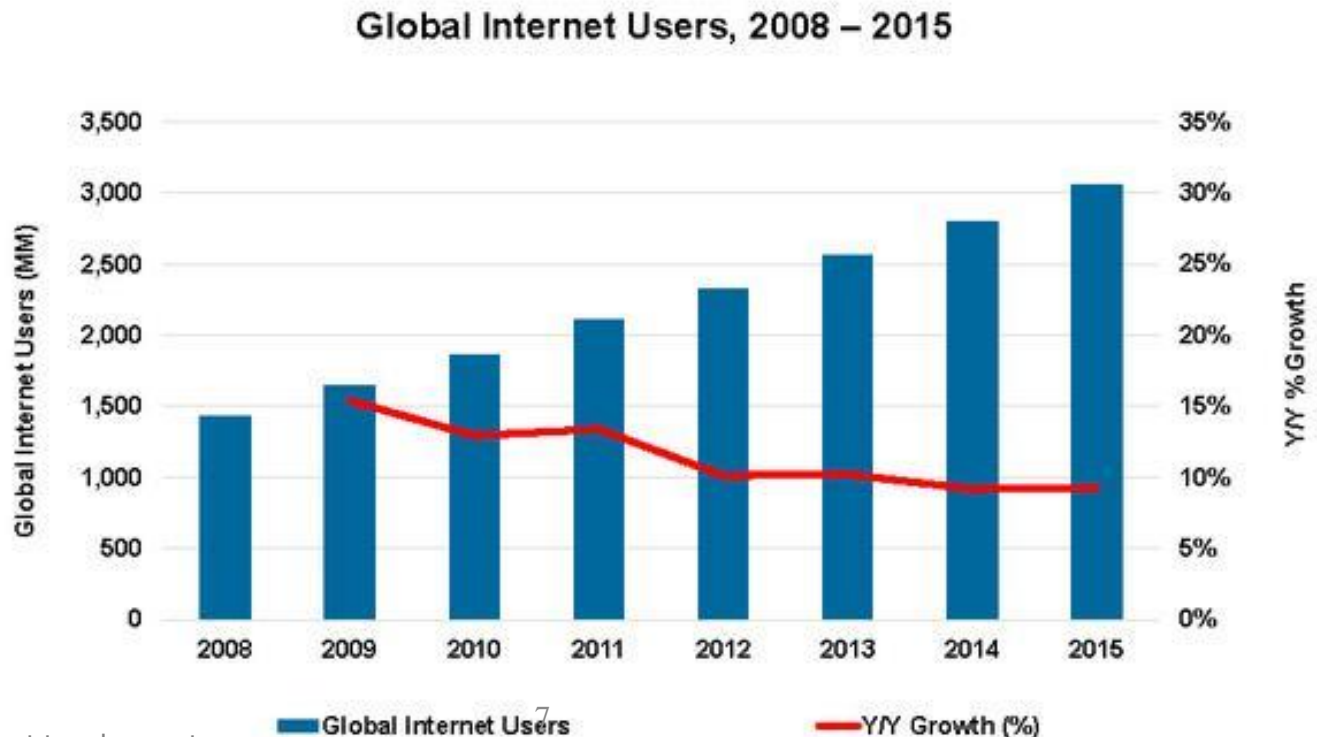


引言：有限资源 \leftrightarrow 无限增长的需求

全球互联网用户

全球互联网用户数已超30亿，比上年增长9%，互联网全球渗透率达到42%。印度互联网用户数达到2.77亿，比上年增长40%，超过美国成为全球第二大互联网市场，仅次于中国。

Global Internet Users = 3B @ 42% Penetration...
+9% vs. +9% Y/Y...+7% (Excluding India)

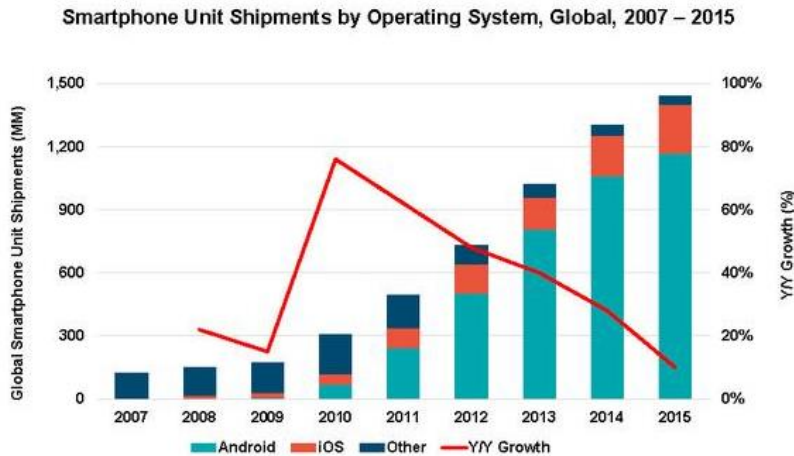




引言：有限资源 \leftrightarrow 无限增长的需求 全球智能手机用户数

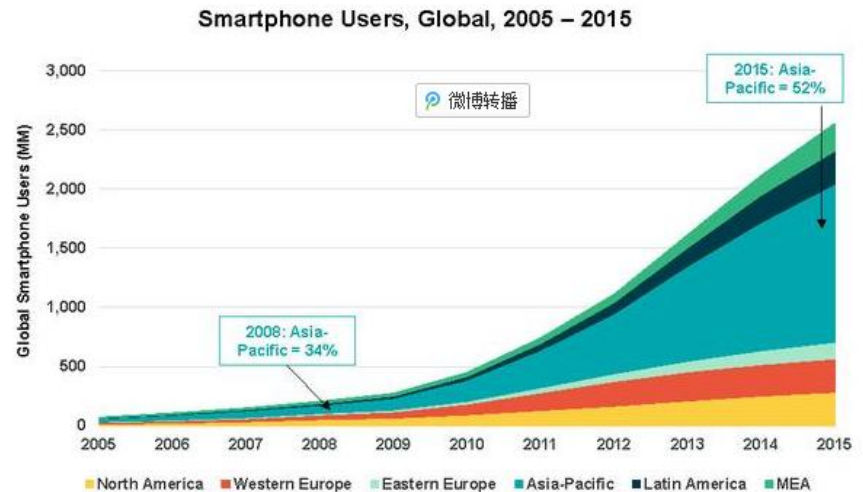
全球智能手机用户数比上年增长21%，去年增长31%，增速放缓；在经过5年高增长后，2015年全球智能手机出货量增长10%，去年增长28%。

Global Smartphone Units Slowing Dramatically...
After 5 Years of High Growth @ +10% vs. +28% Y/Y



Global Smartphone User Growth Slowing...
Largest Market (Asia-Pacific) = +23% vs. +35% Y/Y

亚太地区智能手机用户数增长23%，上年增长35%。





引言：有限资源 \leftrightarrow 无限增长的需求

移动互联网的现状

表 4 2018.12-2019.6 网民各类互联网应用的用户规模及使用率

应用	2019.6		2018.12	
	用户规模(万)	网民使用率	用户规模(万)	网民使用率
即时通信	82470	96.5%	79172	95.0%
搜索引擎	69470	81.3%	68132	82.0%
网络新闻	68587	80.3%	67473	81.0%
网络视频 (含短视频)	75877	88.8%	72486	87.5%
网络购物	63882	74.8%	61011	73.0%
网络支付	63305	74.1%	60040	72.5%
网络音乐	60789	71.1%	57560	70.0%
网络游戏	49356	57.8%	48384	57.0%
网络文学	45454	53.2%	43201	52.0%
旅行预订 ²⁷	41815	48.9%	41001	48.0%
网上订外卖	42118	49.3%	40601	49.0%
网络直播 ²⁸	43322	50.7%	39676	47.9%
网约专车或快车	33915	39.7%	33282	39.0%
网约出租车	33658	39.4%	32988	39.0%
在线教育	23246	27.2%	20123	27.0%

2016.6-2019.6网络直播用户规模及使用率



来源：CNIC 中国互联网络发展状况统计调查

2019.6

2018.6-2019.6网络视频（含短视频）用户规模及使用率



来源：CNIC 中国互联网络发展状况统计调查

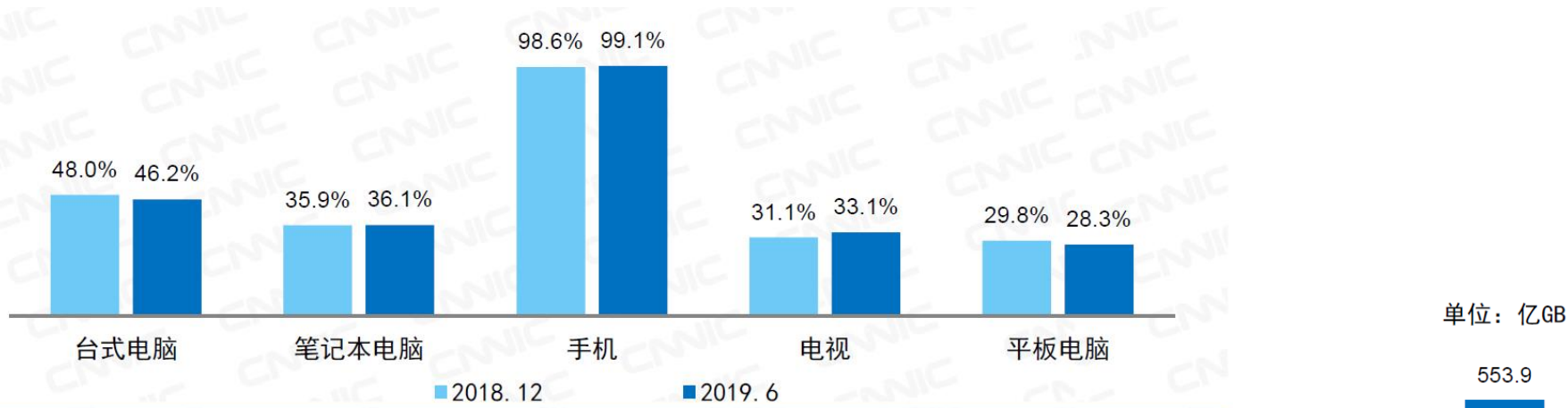
2019.6



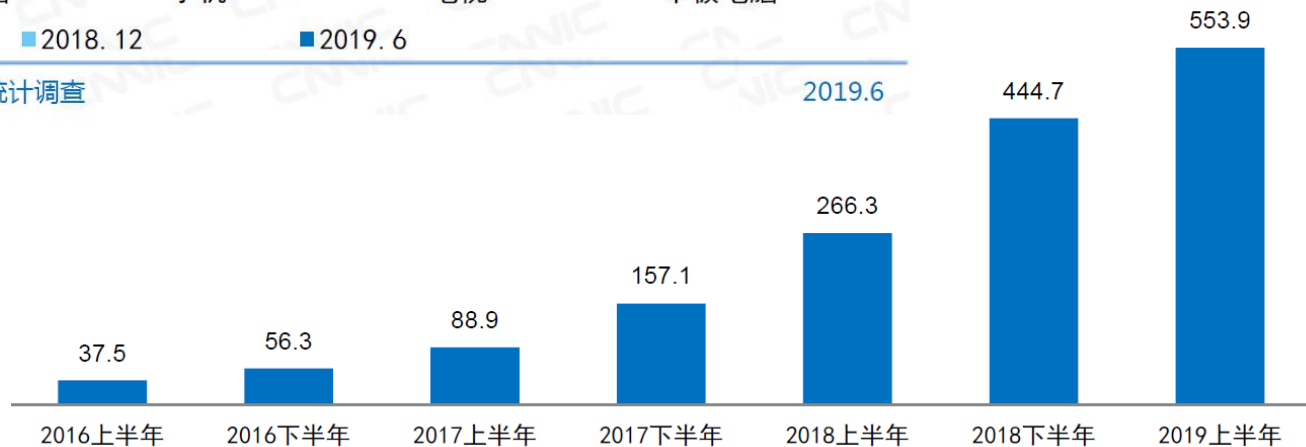
引言：有限资源 \leftrightarrow 无限增长的需求

移动互联网的现状

截止2019年6月，我国网民使用手机上网的比例达99.1%。
2019年1月至6月，移动互联网接入流量消费达553.9亿GB。

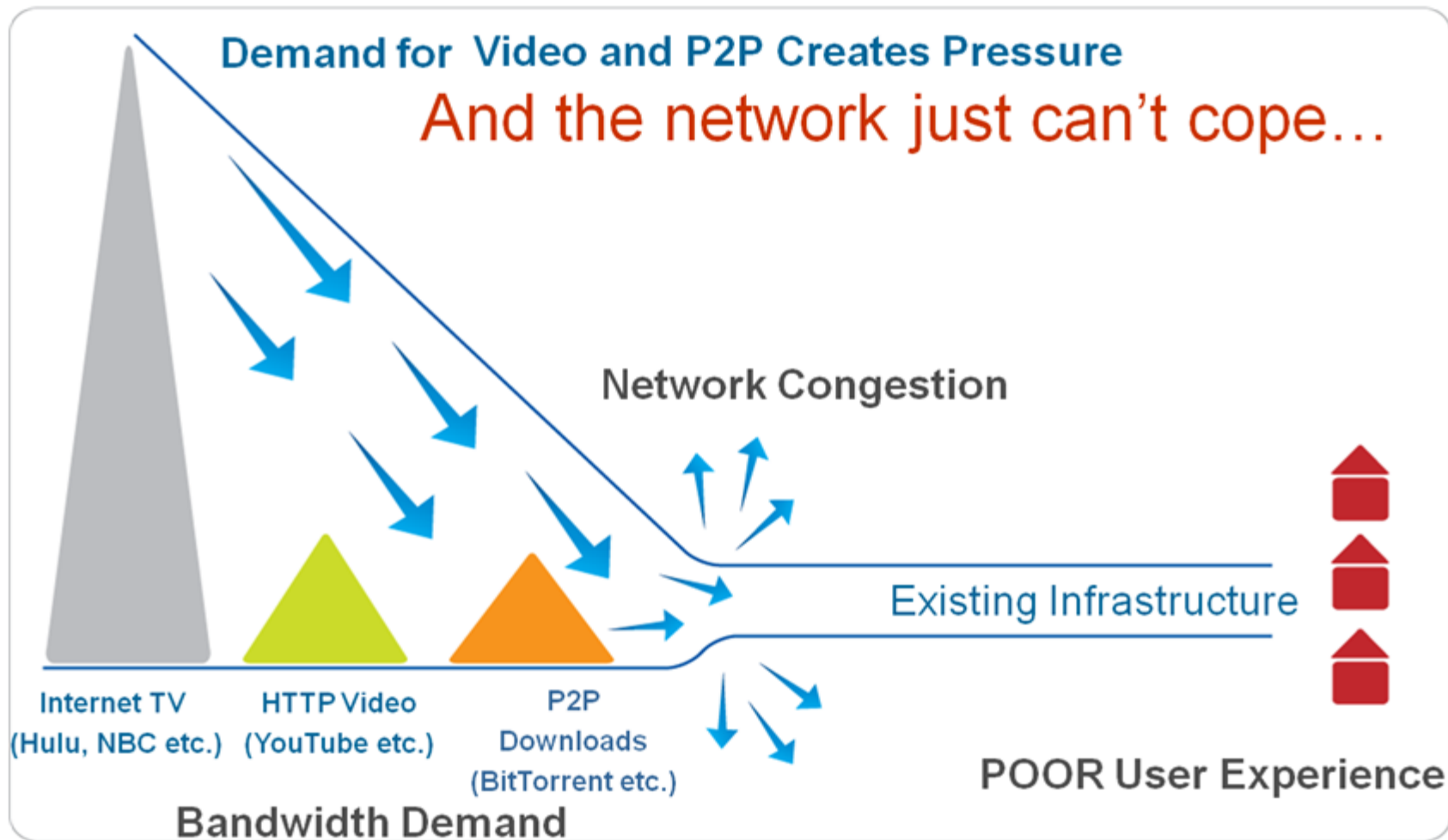


来源：CNIC 中国互联网络发展状况统计调查





引言：有限资源 \leftrightarrow 无限增长的需求 人的需求是永远无法满足的





专题的由来：

感知体验(perceptual experience)的重要性

- ◆ Providing high-quality streaming services over a best effort and shared infrastructure such as the Internet is nontrivial.
- ◆ This is increasingly placing loads on network elements to deliver streaming content with **high perceptual quality**, including content delivery networks (CDNs), overlay networks, video on demand (VoD), and Internet television (IPTV) infrastructures. Network service providers need to **infer, predict, and improve perceptual experience** to ensure long-term success.



专题的由来： 用户体验是根本

- ◆ 衡量一个网络和业务品质的最根本标准在于用户的体验质量
- ◆ QoE, Quality of Experience
 - ITU-T SG12的定义：“the overall acceptability of an application or service, as perceived subjectively by the end user”
- ◆ 用户对于网络的满意度

◆ QoS与QoE概述

- QoS与QoE的概念
- QoS与QoE的区别与联系

◆ 话音质量评估

- 音频的失真
- 话音质量评估方法

◆ 视频质量评估

- 图像和视频的失真
- 图像和视频的质量评估方法
- 网络视频的质量评估



QoS(Quality of Service)参数

◆可用带宽

- 网络的两个节点之间特定业务流的平均速率

◆时延

- 数据包在网络的两个节点之间传送的平均往返时间

◆丢包率

- 在网络传输过程中丢失报文的百分比

◆时延抖动

- 时延的变化

◆误包率

- 网络传输中报文出现错误的百分比。



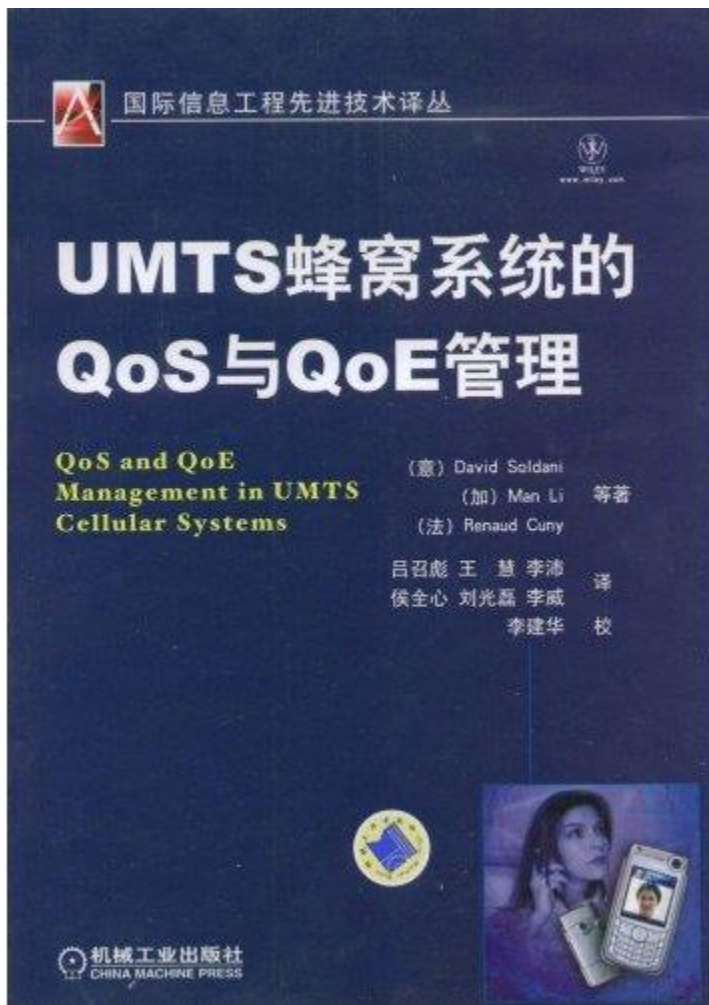
QoE (Quality of Experience)

- ◆ QoE是用户对网络服务的满意程度
- ◆ QoS是一个技术概念，是通过网络或网络元素来测量、表示，但对用户而言通常没有多大的意义。
- ◆ QoE定义为一个应用或业务的总体可接受性，是终端用户的主观感知。如网络业务用户认为应该达到的页面显示或文件的传送速度、音频质量、图像质量、视频连续度、音 / 视同步程度等。

MC QoE的通俗表示

- ◆ QoE可以用5个级别定性界定
 - 优秀(excellent: imperceptible)
 - 好(good: perceptible, but not annoying)
 - 一般(average: slightly annoying)
 - 差(bad: annoying)
 - 极差(terrible: very annoying)

MC 参考书





VoIP业务QoE等级与QoS参数关系

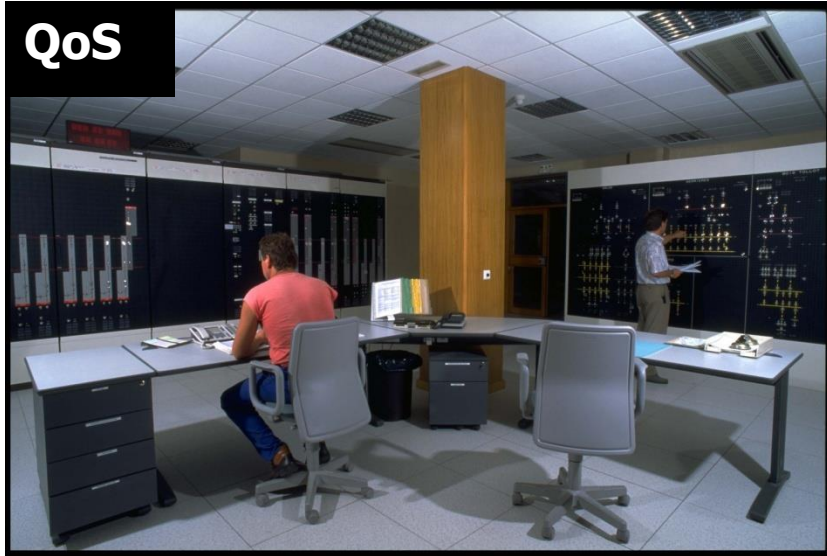
VoIP 业务 QoE 等级	用户主观感觉	延时 (ms)	抖动 (ms)
优秀	声音清晰, 效果好	<100	<20
好	感觉不到延时和抖动	100 - 150	20 - 40
一般	可以接受, 但偶尔可以感觉到微小的延时和抖动	150 - 250	40 - 75
差	不能接收, 很多杂音	250 - 300	75 - 90
极差	不能进行通话	>300	>90



常见多媒体业务对传输网络的要求

要求	最大延迟 /s	最大延迟 抖动/ms	速率 /Mbps	平均吞吐 量/Mbps	可接受 误码率	可接受分 组差错率
话音	0.25	10	0.005-0.064	0.064	$<10^{-2}$	$<10^{-2}$
视频图像	0.25	10	100-1000	100	10^{-2}	10^{-3}
静止图像	1	/	2-10	2-10	10^{-4}	10^{-9}
压缩后视频 图像	0.25	1	0.02-10	2-10	10^{-6}	10^{-9}
数据文件	1	/	2-100	2-100	0	0
实时数据	0.001-1	/	<10	<10	0	0

MC QoS vs. QoE



■ Quality of Service

- ◆ Network-centric
- ◆ Delay, packet loss, jitter
- ◆ Transmission quality
- ◆ Content agnostic

■ Quality of Experience

- ◆ Content impairments
- ◆ Blockiness, Jerkiness, ...
- ◆ End-user quality
- ◆ Application driven



例：同样的QoS不同的QoE



Same network impairments

Packet Loss: 1%

Delay: 10ms

Jitter: 50us

Bandwidth: 500kbps

Different perceived quality!

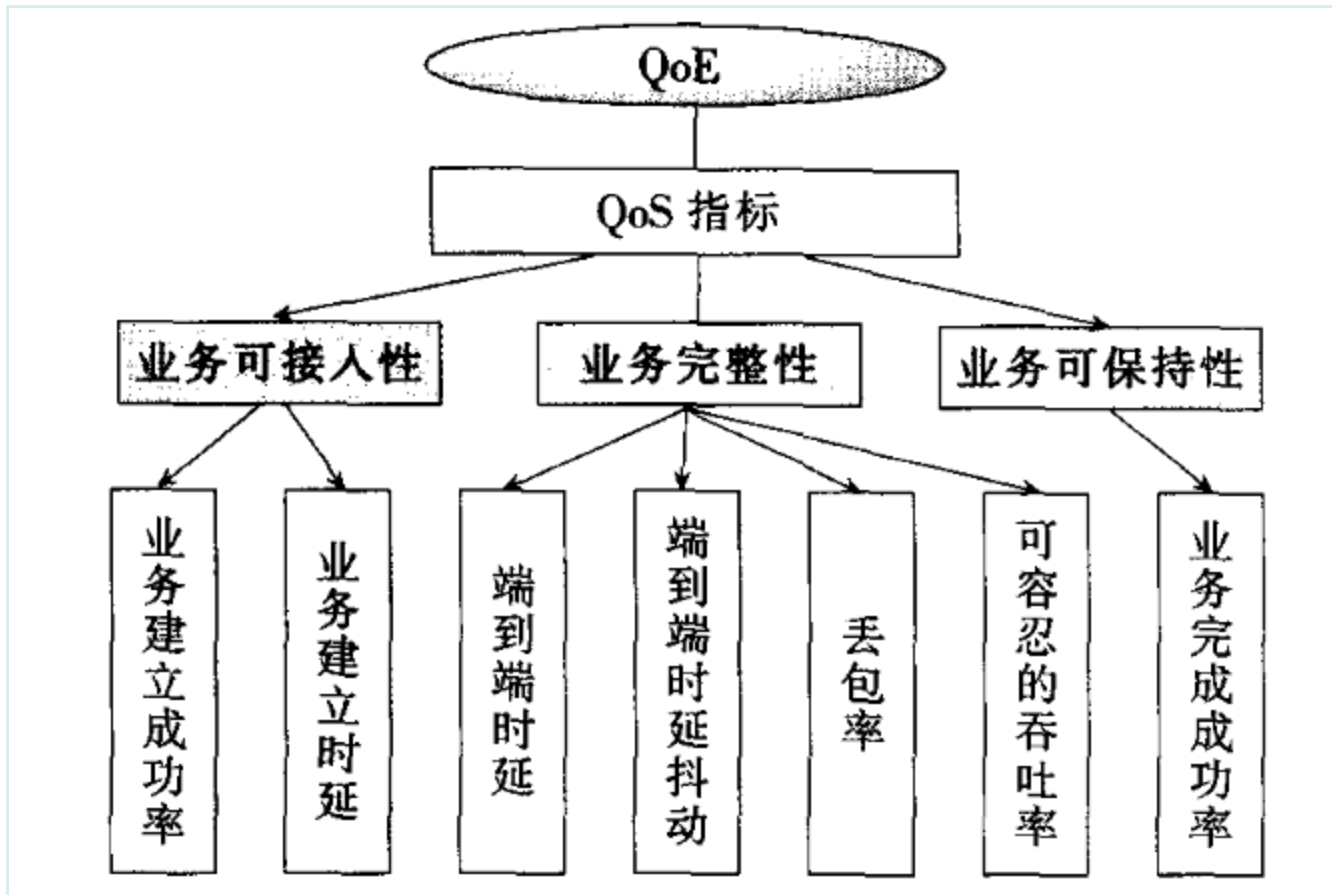


QoS不足以反映用户感受

- ◆ 现有网络中没有基于**全网**的性能监控指标，同时现有QoS体系与用户真实体验存在距离，单纯的分段式的网络性能分析已经不能满足网络优化需要。
- ◆ 系统化的、端到端的用户体验质量(QoE)、QoS优化分析成为业界研究的热点问题。



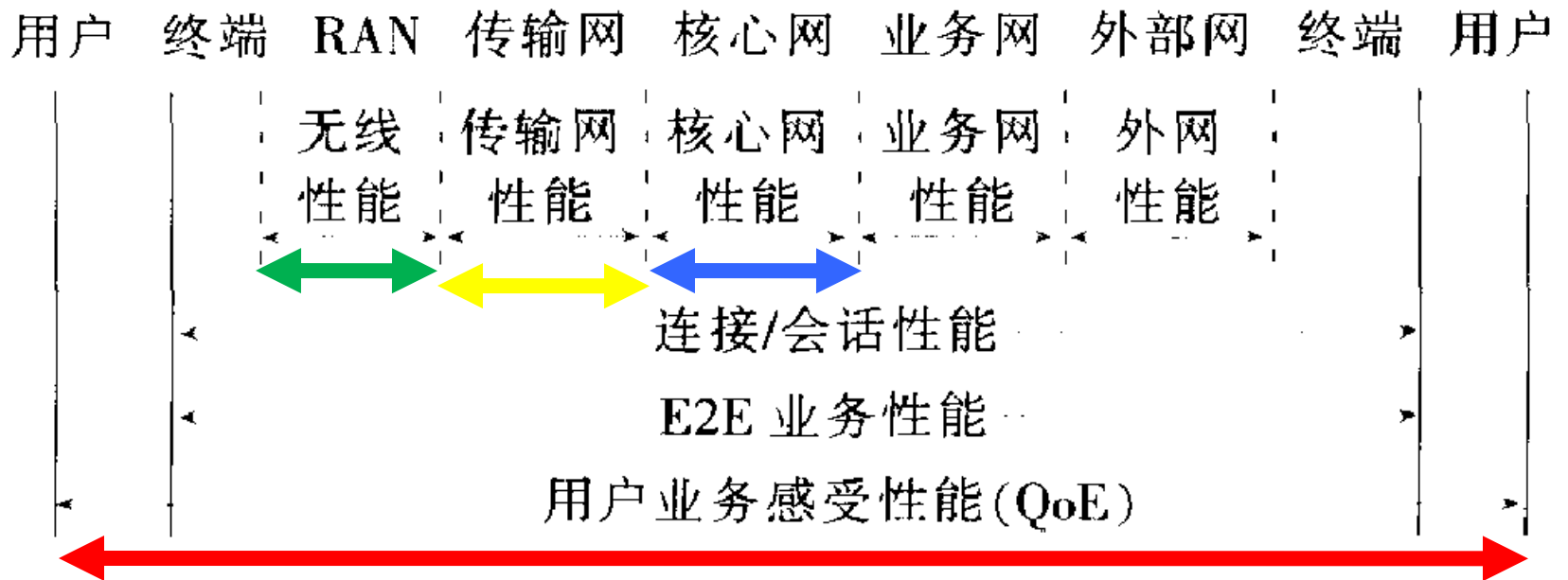
基于QoS指标进行QoE评估?





QoE=网络QoS+内容+人的体验

- ◆ QoS在具体的通信网络中体现
- ◆ QoE和具体的业务紧密联系

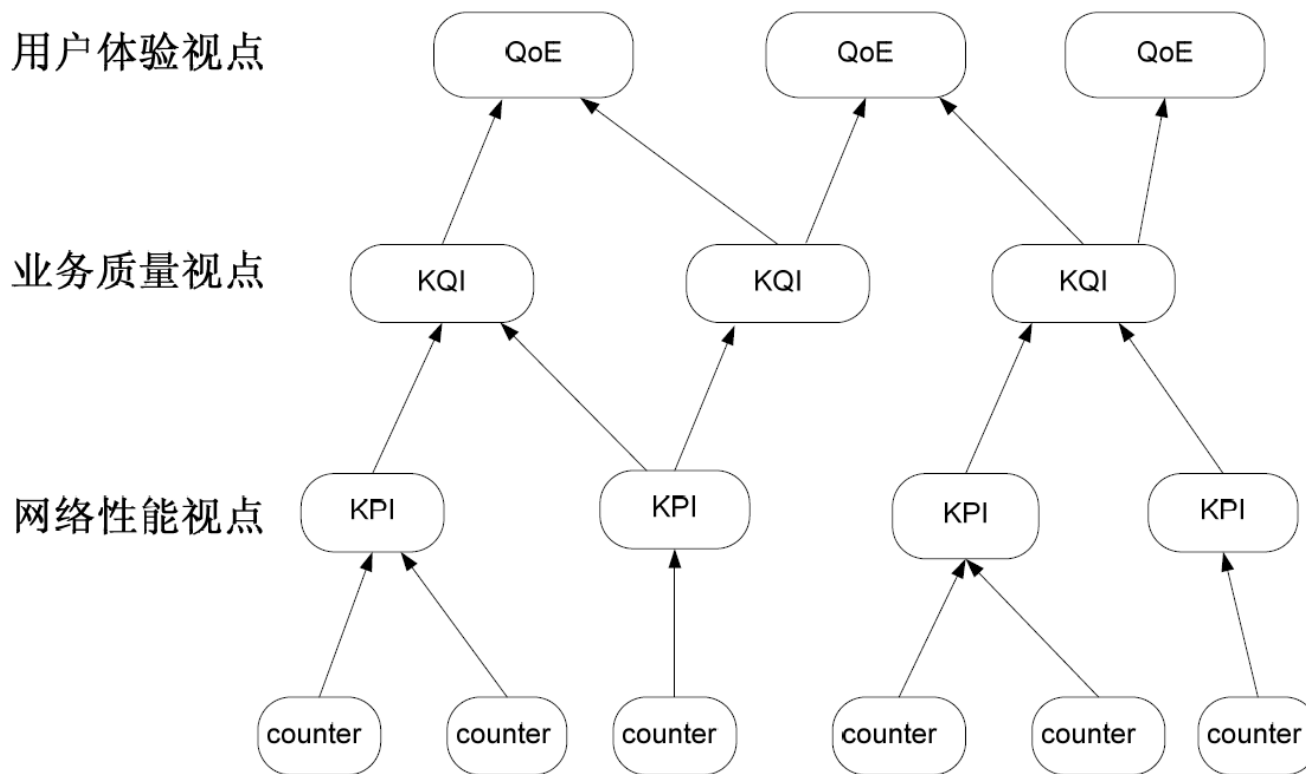




技术层面的监控：

QoE、KQI、KPI和网元设备counter之间的关系

- ◆ KPI (Key Performance Indicator, 关键性能指标)
- ◆ KQI (Key Quality Indicator, 关键质量指标)



MC 小结： QoS与QoE

◆ 科技以人为本

- WEB2.0/通信生态系统/贪婪的需求

◆ QoE以人的感受为价值核心

- QoS是技术平面的问题
- QoE是用户平面的问题

◆ QoE重在管理

- 与QoS在技术平面的映射
- 用户满意度？

- ◆ QoS与QoE概述
- ◆ 话音质量评估
 - 音频的失真
 - 模拟音频的失真
 - 数字音频的失真
 - 电话话音的失真
 - 话音质量评估方法
 - PESQ评估：计算MOS
 - E-Model评估：计算R因子
- ◆ 视频质量评估
 - 图像和视频的失真
 - 图像和视频的质量评估方法
 - 网络视频的质量评估



模拟音频的失真：谐波失真

- ◆ 由于电路的非线性，导致输出信号产生了输入信号所没有的谐波分量即为总谐波失真。
- ◆ 在总谐波失真中，相对来说，偶次谐波对基频信号的不良影响要比奇次谐波小得多。偶次谐波的出现虽然会影响基频的清晰度，但它却会为基频添加“温暖”、“厚实”及“明亮”的感觉；因为从音律学的角度看，对基频影响较大的偶次谐波中的2次、4次谐波与基频形成的是协和的八度音程关系。因此当总谐波失真中的主要成分是偶次分量时，人们对它所带来的失真的感知并不明显，相反有时甚至觉得这种失真的存在是自然和正常的。而奇次谐波给人的感觉是“缺乏层次”和“生硬的毛刺感”，因为奇次谐波尤其是7次以上的奇次谐波，与基频形成了不协和的音程关系，干扰并损坏了基频的音色。



模拟音频的失真：互调失真

- ◆ 由于电路的非线性，导致两个以上的输入信号在放大器的输出端产生了频率或幅度调制，即为互调失真。
- ◆ 例：假设有两个频率为300Hz和500Hz的纯音信号，当它们通过放大器的非线性区后，在输出端将不但会出现它们各自的谐波，而且还会产生800Hz的和频及200Hz的差频信号。由于这个和频及差频均不能与这两个基频构成协和的音程关系（非大三或小三也非纯五或八度音程），所以它们的出现是有损于基频的。不仅如此，如果电路的非线性越差，频率调制的程度就越高，这时不仅存在着两个基频之间的调制，它们各自的谐波之间、谐波与基频之间也存在着相互的频率调制。正是这些和频与差频及其高次频率调制分量的集合，大大改变了原有音色。



模拟音频的失真：噪声

- ◆ 噪声包括：随机噪声和相干噪声。随机噪声在此主要是指器件的热噪声（有源器件也称散粒噪声）；相干噪声则主要是感生噪声，例如电源的感生交流声、射频干扰等。
- ◆ 高斯噪声或白噪声也就是常说的热噪声，由于它是因器件中电子的热运动引起的，所以能量在整个频谱上是均匀分布的。举例来说，虽然从八度关系上看，100Hz与200Hz都各是一个8度，但后者所包含的热噪声能量却是前者的100倍。不难看出，就热噪声而言，仅有5%的能量存在于20Hz-1KHz之间，而95%的能量是在1kHz-20kHz之间的，其中10kHz-20kHz中的热噪声能量又占50%。因此，热噪声在表现上给人以“嘶嘶”的高频噪声感觉。



数字音频的失真：量化失真

$$\begin{aligned} SNR &= 10 \lg [(V_{\text{signal}})^2 / (V_{\text{noise}})^2] \\ &= 20 \lg (V_{\text{signal}} / V_{\text{noise}}) \end{aligned}$$

其中：

V_{signal} 表示信号电压；

V_{noise} 表示噪声电压；

SNR的单位为分贝(db)。

例1：假设 $V_{\text{noise}}=1$ ，量化精度为1比特表示 $V_{\text{signal}}=2^1$ ，它的信噪比 $SNR=6$ 分贝。

例2：假设 $V_{\text{noise}}=1$ ，量化精度为16比特表示 $V_{\text{signal}}=2^{16}$ ，它的信噪比 $SNR=96$ 分贝。



数字音频的失真：编码失真

- ◆ MOS(Mean Opinion Score)
- ◆ 常见编码器的MOS分值如下

Coder	MOS
Quiet room 48 KHz 16 bit linear sampling	5.0
PCM (A-law/mlaw) 64 Kb/s	4.1
G.723.1 @ 6.3 Kb/s	3.9
G.729 @ 8 Kb/s	3.9
ADPCM G.726 32 Kb/s	3.8
toll quality 【与PSTN一样的质量】	
GSM @ 13Kb/s	3.6
VSELP IS54 @ 8Kb/s	3.4

MC 音频→电话话音

◆ 电话话音与一般音频不同

- CD-DA、电视等是单向播放的，电话话音是交互式的
- 电话话音要经过电话网络的传输

◆ 电话语音质量对于大多数情况来说是一个主观因素，对于**清晰**、**低时延**、**低抖动**的话音我们认为其质量是可以接受的。由人对话音的质量直接评估是语音质量的主观评价；也有可以量化的语音质量评判标准。

PSTN(Public Switched Telephone Network)



可感知的电话话音失真：时延

- ◆ ITU G.114规范建议，在传输语音流量时，单向语音包**端到端延迟**要低于150ms(对于国际长途呼叫,特别是卫星传输时,可接受的单向延迟为300ms。如果**超过300ms**则通话的质量会变的让人不能忍受。过多的包延迟可以引起通话声音不清晰、不连贯或破碎。
- ◆ 大多数用户**察觉不到小于100毫秒的延迟**，当延迟在100毫秒和300毫秒之间时，说话者可以察觉到对方回复的轻微停顿。这种停顿可以影响到通话双方的交流。超过300毫秒，延迟就会很明显，用户开始互相等待对方的回复，通话过程变成类似对讲机式的模式。



ITU G.114

One-way transmission time

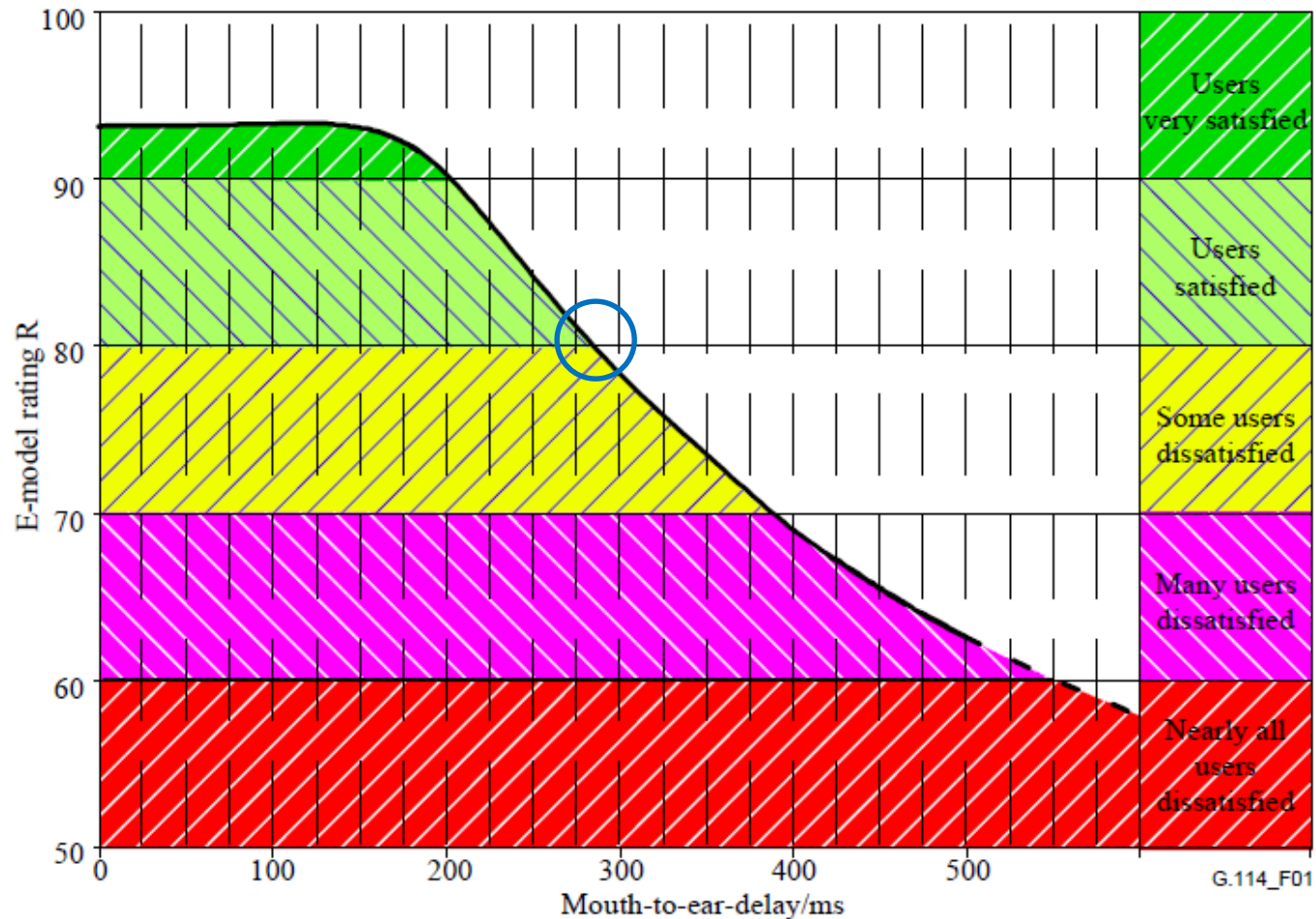


Figure 1/G.114 – Determination of the effects of absolute delay by the E-model



可感知的电话话音失真：回声

- ◆ 回声（Echo）一般分为说话者的回声和倾听者的回声。说话者的回声就是在通话过程中，说话者听到了自己的声音。倾听者的回声就是倾听者重复听到说话者的声音。
- ◆ 回声产生的时间延时较小时，给人的感觉只是简单适度的侧音（就像我们平常说话时，我们通过颞骨听到自己的声音），这样的侧音对我们来说是习惯的。但当回声延迟超过25毫秒时，回声不利的影响就会被说话者察觉。



可感知的电话话音失真：沿切割

- ◆ 丢失说话人的第一个或最后一个音节或单词的现象叫沿切割（Clipping）。包括“前沿切割”（Leading Edge Clipping）和“后沿切割”（Trailing Edge Clipping）。
- ◆ 沿切割由延迟和IP网络中的静音抑制机制的使用而引起。为了更有效地使用网络带宽，常使用静音检测机制，例如静音检测（VAD）或者舒适噪音生成器（CNG）。静音检测器(VAD = Voice Activation Detector) 在信号电平低于某一特定的门限值时，将限制数据包的传输，此时提供空闲（Idle）或者“舒适噪音”（Comfort Noise）以避免电话用户感到“断线”（Dead Air）。从静音→讲话和从讲话→静音之间的过渡状态的响应速度如果过慢，就会丢失开始或最后的音节。



可感知的电话话音失真类型

◆ 话音失真现象往往不是单一产生的，时常是多种失真伴随出现，比如说时延，抖动，和沿切割。

- 话音编解码 (Encoding and Decoding of Voice)
- 断续 (Time-Clipping, 也称为Dropouts)
- 延时抖动(Jitter)
- 环境噪音(Environment Noise)
- 信号衰减(Signal Attenuation)
- 传输信道误码(Transmission Channel Errors)



话音评估的主要指标分类

Primary VoIP Quality Metrics

◆ 电话话音的质量包括

- 话音清晰度 (Voice Clarity)
- 回声 (Echo)
- 时延 (Time Delay)

回声和时延可以精确到毫秒级的客观指标，在实际评估方面不会有何争议；相反，人们对于清晰度却是仁者见仁，智者见智，缺乏一种统一的评判标准。

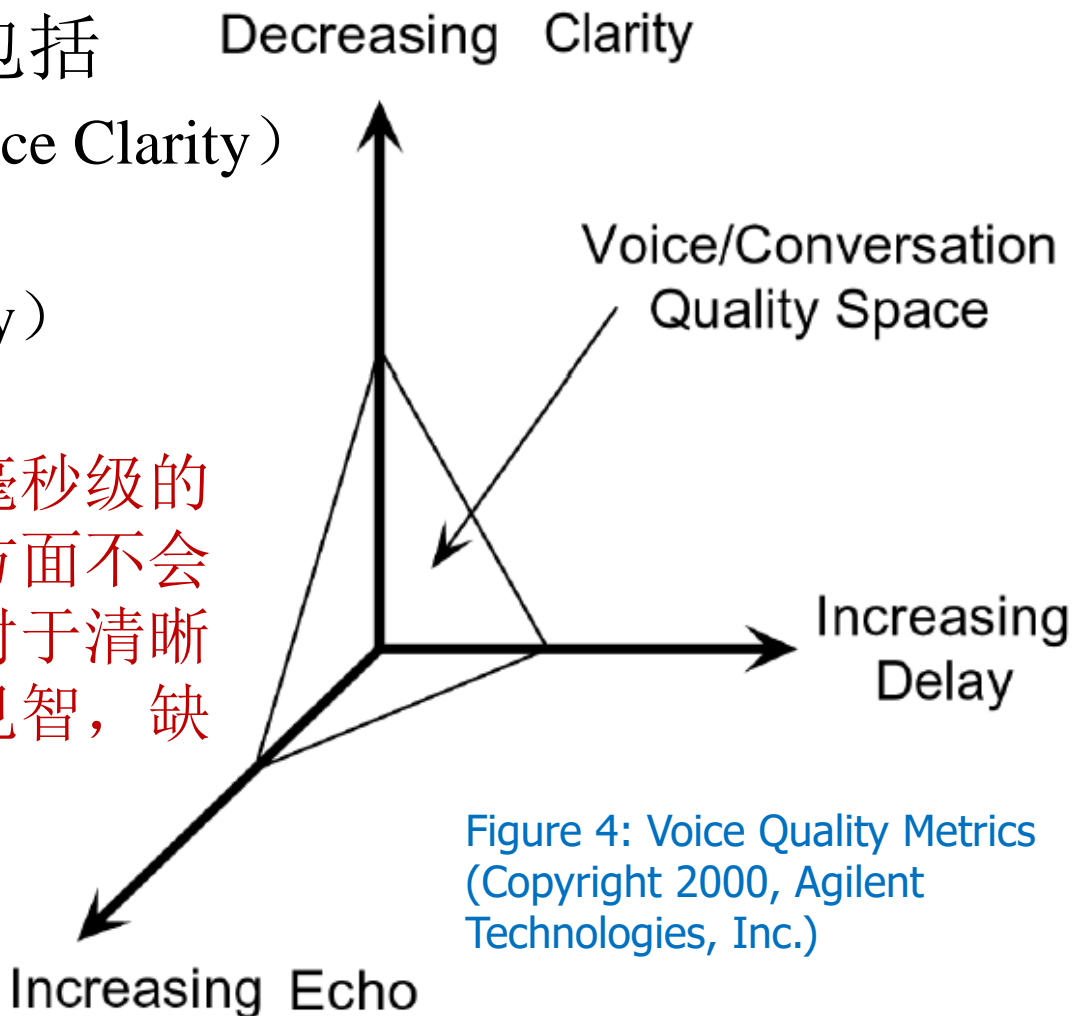


Figure 4: Voice Quality Metrics
(Copyright 2000, Agilent Technologies, Inc.)

◆ QoS与QoE概述

◆ 话音质量评估

□ 音频的失真

- 模拟音频的失真
- 数字音频的失真
- 电话话音的失真

□ 话音质量评估方法

- PESQ评估：计算MOS
- E-Model评估：计算R因子

◆ 视频质量评估

□ 图像和视频的失真

□ 图像和视频的质量评估方法

□ 网络视频的质量评估



语音质量的评估方法

◆ 主观方法

- MOS-LQ : Mean Opinion Score-Listening Quality
- DMOS: 参考信号和被评估信号MOS值之差
- 主观的评价只能通过调查获取，但人的感受受到很多因素的影响：IQ、EQ、心理、疲劳程度...

◆ 客观方法

- ITU组织推行了大量的端到端语音质量客观测试技术的标准化工作，发布了几种语音评估算法标准：PAMS、PSQM、PSQM+、MNB、PESQ。MOS评测开始摆脱原始的主观评估方式，而**使用与量化算法计算相对应的级别**及语音质量好坏程度。



常用话音质量评估方法分类

◆ 基于输入-输出方式的客观评价方法

- 又称为侵入式(Intrusive)语音评价。通过建立人耳听觉模型，并对原始语音信号和失真语音信号进行误差对比来获得相应的参数，再根据参数和语音质量评价分数之间的映射关系得到最终的评价结果。
- PSQM (Perceptual Speech Quality Measurement), ITU-T P.861, 1996
- PAMS (Perceptual Analysis Measurement System)
- PESQ (Perceptual Evaluation of Speech Quality), ITU-T P.862, 2001

◆ 基于输出的客观评价方法

- 又称为非侵入式(Non-Intrusive)话音评价。仅通过语音系统的输出信号和影响语音系统性能的参数来评估语音质量的好坏,不需要原始参考信号的参与,便于进行系统的实时话音质量里评估。由于通话环境比较复杂,和输入-输出方法相比这类算法的评估效果有所下降。
- E-Model, ITU-T的G.107, 1998

ITU-T G.107, <http://www.itu.int/rec/T-REC-G.107/en>

The E-model: a computational model for use in transmission planning

ITU-T P.861, <http://www.itu.int/rec/T-REC-P.861/en>

Objective quality measurement of telephone-band (300-3400 Hz) speech codecs

ITU-T P.862, <http://www.itu.int/rec/T-REC-P.862/en>

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of [yinh, cxh](mailto:yinh_cxh@ustc.edu.cn) @ustc.edu.cn

networks and speech codecs



ITU 关于语音评估的标准

- ◆ P.563, Single-ended method for objective speech quality assessment in narrow-band telephony applications
- ◆ P.800 Methods for subjective determination of transmission quality
- ◆ P.800.1 Mean Opinion Score (MOS) terminology
- ◆ P.861, Objective quality measurement of telephone-band (300-3400 Hz) speech codecs
- ◆ **P.862 Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs**
- ◆ P.862.1 Mapping function for transforming P.862 raw result scores to MOS-LQO
- ◆ P.862.2 Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs
- ◆ **G.107, E-Model, a computational model for use in transmission planning**



PESQ的基本思路

Perceptual evaluation of speech quality (PESQ)

◆ 已应用的测量方法

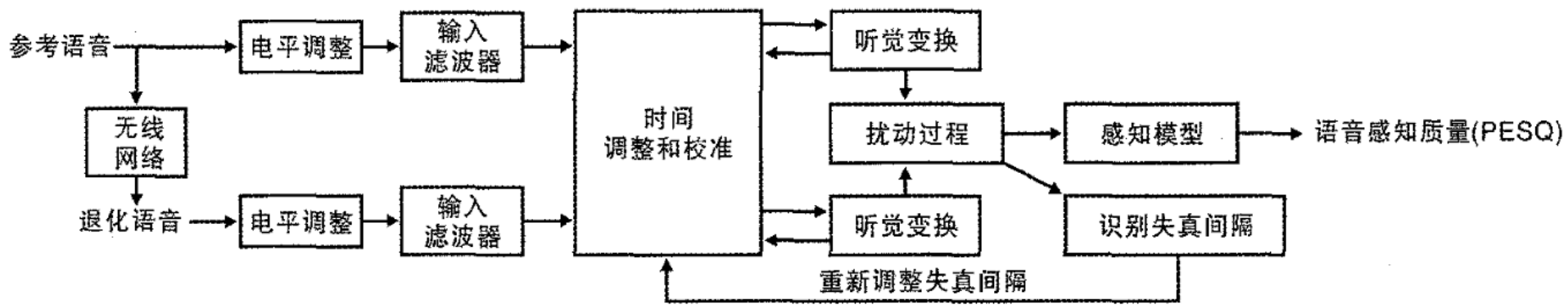
- PSQM /PSQM+感知通话质量测量
- PESQ感知评估通话质量测量
- PAMS（英国电信）感知分析测量等

◆ 测量方法的基本思路：需要发送一个语音参考信号通过电话网络，在网络的另一端采用数字信号处理的方式比较样本信号和接收到的信号，进而估算出网络的语音质量。

◆ PESQ结合了PSQM和PAMS的优势，针对VoIP和混合的端到端应用作了改进，并针对MOS和MOS-LQ计算方法做了修改。

MC PESQ流程

◆参考信号和通过网络传输后的信号通过电平调整，再用输入滤波器模拟标准电话听筒进行滤波(FFT)。这两个信号在时间上对准，并通过听觉变换。这个变换包括对系统中线性滤波和增益变化的补偿和均衡，再通过认知模型，从而映射到对主观平均意见分的预测。一般情况下，输出信号和参照信号的差异性越大，计算出的MOS分值就越低。



ITU-T P.862, <http://www.itu.int/rec/T-REC-P.862/en>

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

{ynh, cxh}@ustc.edu.cn



丢包率与PCM MOS关系

		Subjective MOS-LQO	PESQ MOS-LQO	Error
3% Packet Loss	G.711 no PLC	3.23	3.14	0.09
	G.711 Appendix I	4.02	3.75	0.27
	G.711 + NetEQ	4.18	3.26	0.92
5% Packet Loss	G.711 no PLC	2.74	2.94	-0.20
	G.711 Appendix I	3.80	3.48	0.32
	G.711 + NetEQ	4.04	3.25	0.79
10% Packet Loss	G.711 no PLC	1.99	1.93	0.06
	G.711 Appendix I	3.50	2.58	0.92
	G.711 + NetEQ	3.83	2.34	1.49



VoLTE语音质量MOS

The overall quality target for VoLTE is for voice calls to be perceived as on a par with, or better than, a 3G call. Using WCDMA as the 3G benchmark, the default target is for speech quality to be greater than 3.5 MOS-LQSM on a call-by-call basis, as illustrated in **Figure 1**.

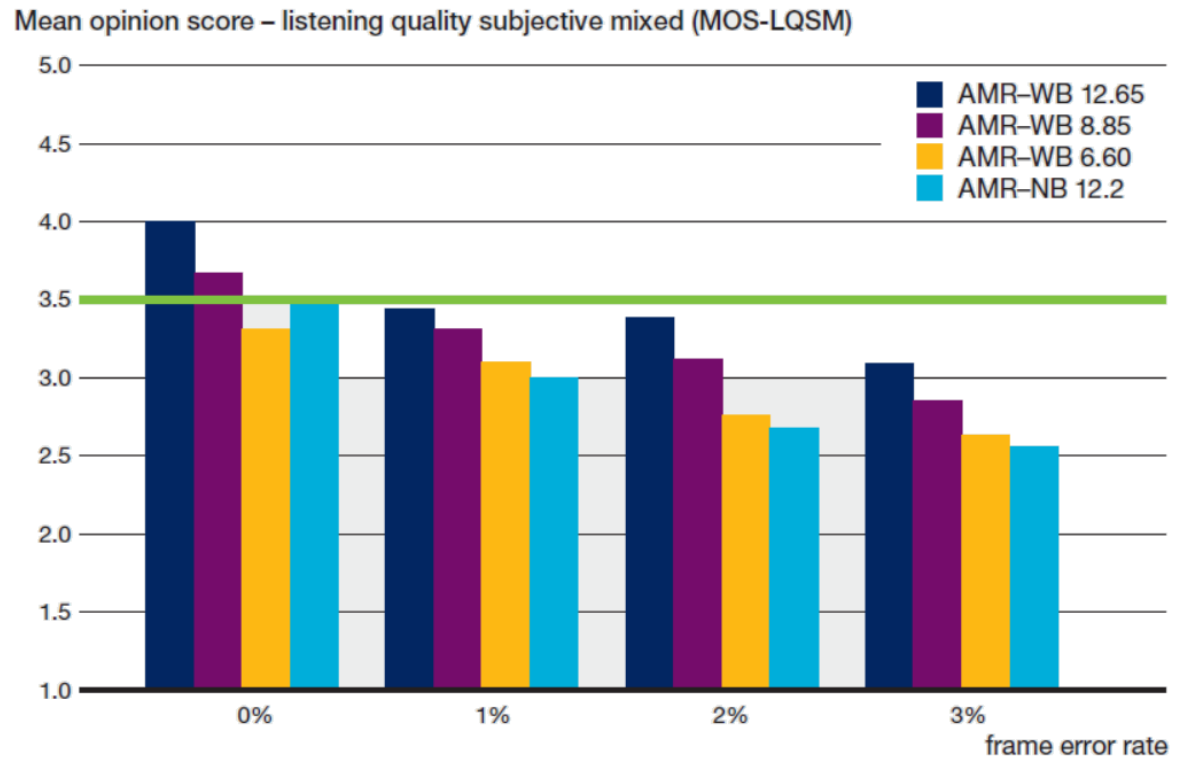
VoIP: voice over IP

VoLTE: voice over LTE

MOS: mean opinion score

MOS-LSQM: mean opinion score – listening quality subjective mixed

FIGURE 1 Speech quality versus frame error rate

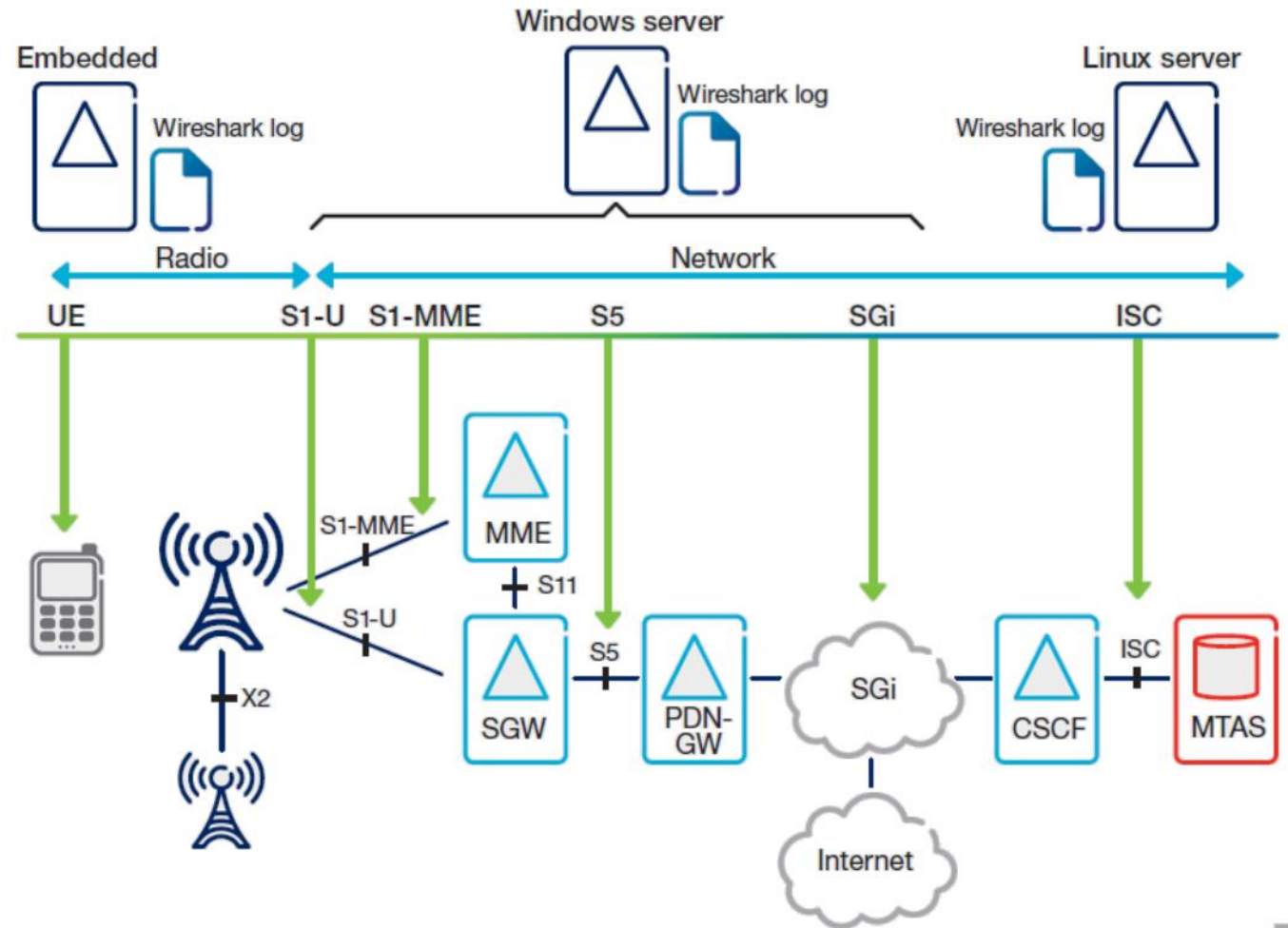




VoLTE语音测试数据采集点

Over-the-air test system, indicating that the UE is connected to the lab network via the air interface, not via a coaxial cable.

FIGURE 4 Data capture points



MC E-model 动机

- ◆ PSQM、PAMS、PESQ等方法通过比较收发信号之间的差异来进行话音质量的评估。不能反映诸如延时、抖动和丢包等数据网络特有的问题。
- ◆ 为了克服这些缺点，ITU G.107标准提出了E-model话音质量评估模型，先根据网络延迟、噪声以及话音编解码器等环境参数计算出传输性能等级系统R，再将其转化为MOS。从而可以关注到网络损伤因素。
- ◆ The E-model is based on the equipment impairment factor method, following previous transmission rating models. It was developed by an ETSI ad hoc group called "Voice Transmission Quality from Mouth to Ear".



E-Model思想

◆ 算法输出R作为评估因子

□ rating factor, 取值0到100

$$R = R_o - I_s - I_d - I_e + A$$

◆ R_o 无网络延时和设备损伤时的基本信噪比

◆ I_s 与语音信号传输同步的损伤

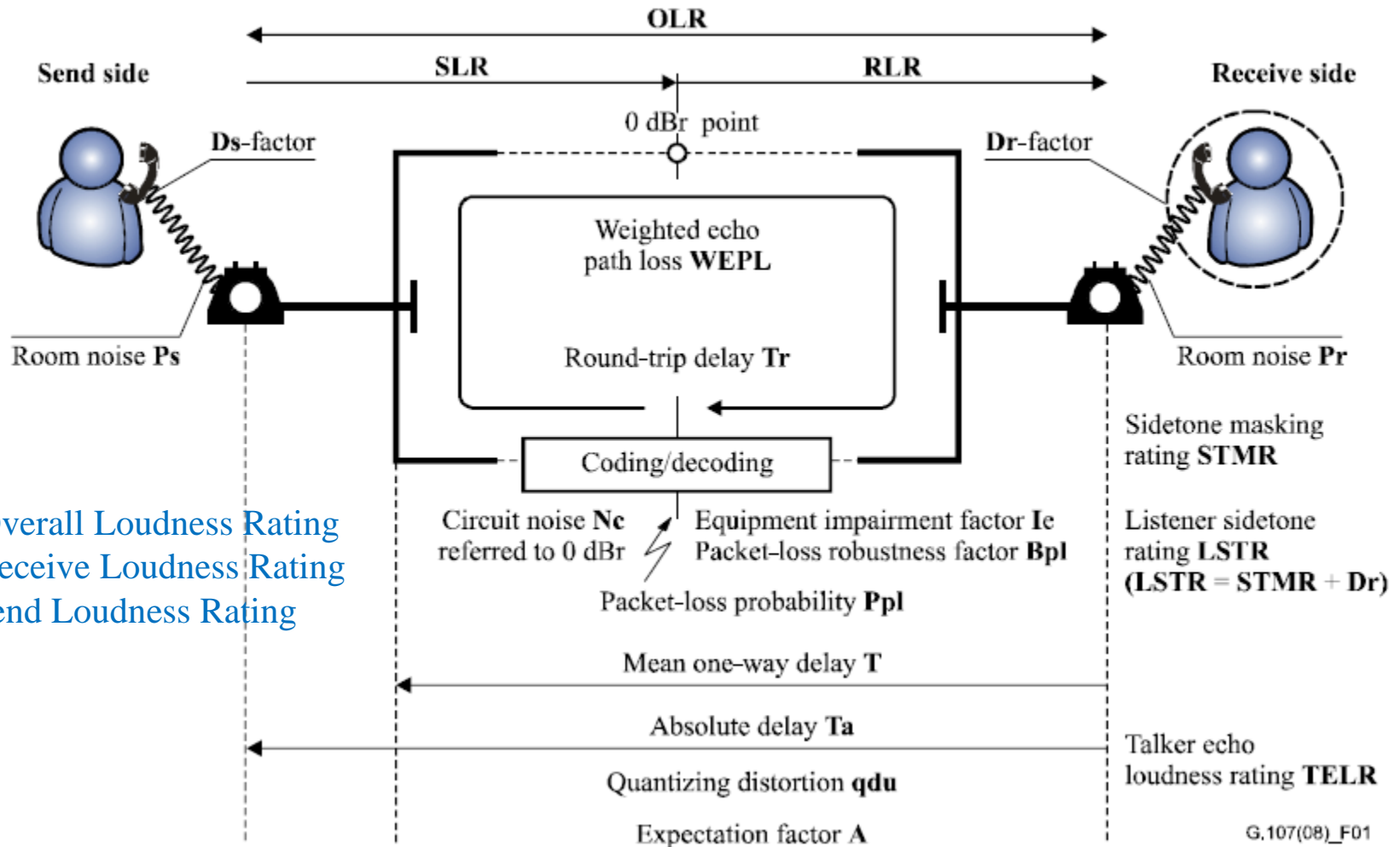
◆ I_d 语音信号传输延时后的损伤

◆ I_e 由设备引入的损伤如编码器损伤

◆ A 优势因素, 考虑呼叫者的期望因素, 一般设置为0



Reference connection of the E-model



OLR Overall Loudness Rating
 RLR Receive Loudness Rating
 SLR Send Loudness Rating

G.107(08)_F01



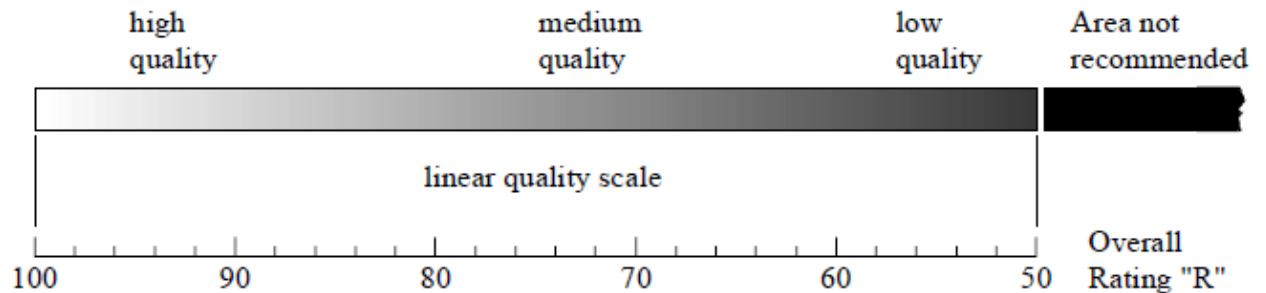
ITU G.109对音频传输质量的分类

Table 1/G.109 – Definition of categories of speech transmission quality

R-value range	Speech transmission quality category	User satisfaction
$90 \leq R < 100$	Best	Very satisfied
$80 \leq R < 90$	High	Satisfied
$70 \leq R < 80$	Medium	Some users dissatisfied
$60 \leq R < 70$	Low	Many users dissatisfied
$50 \leq R < 60$	Poor	Nearly all users dissatisfied

NOTE 1 – Connections with R-values below 50 are not recommended.

NOTE 2 – Although the trend in transmission planning is to use R-values, equations to convert R-values into other metrics e.g. MOS, %GoB, %PoW, can be found in Annex B/G.107.

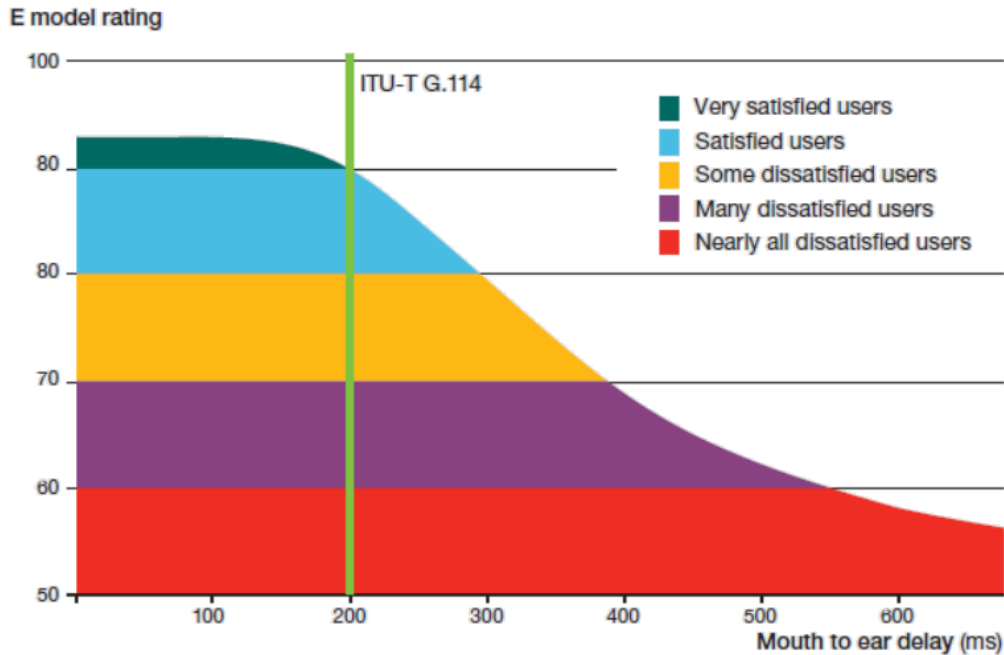


T1211030-99



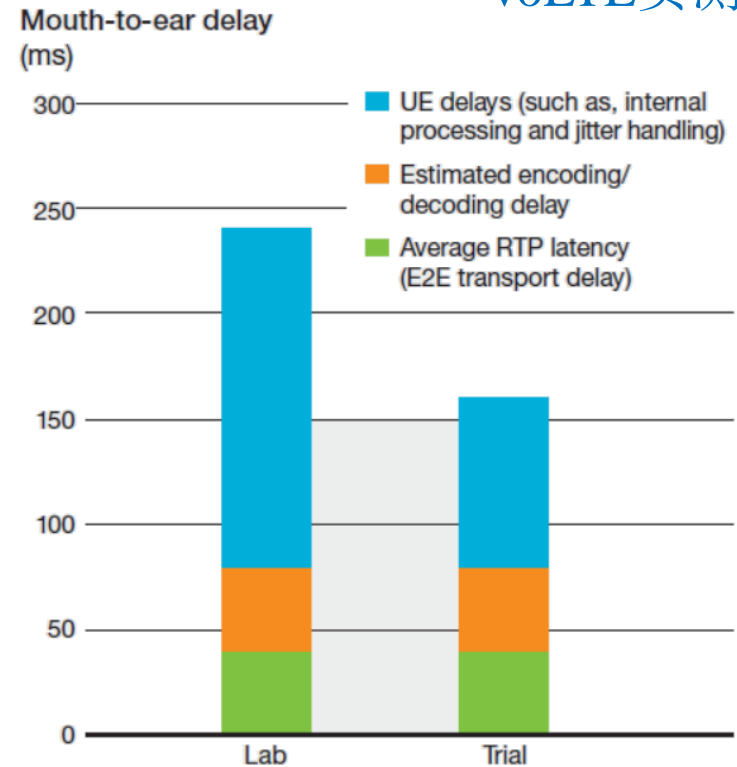
Speech quality versus mouth-to-ear delay

The focus here is the network, so in terms of transport parameters, this means a frame error rate (FER) of less than 1 percent and a mouth-to-ear delay of less than 200ms (reflecting a UE to UE call).



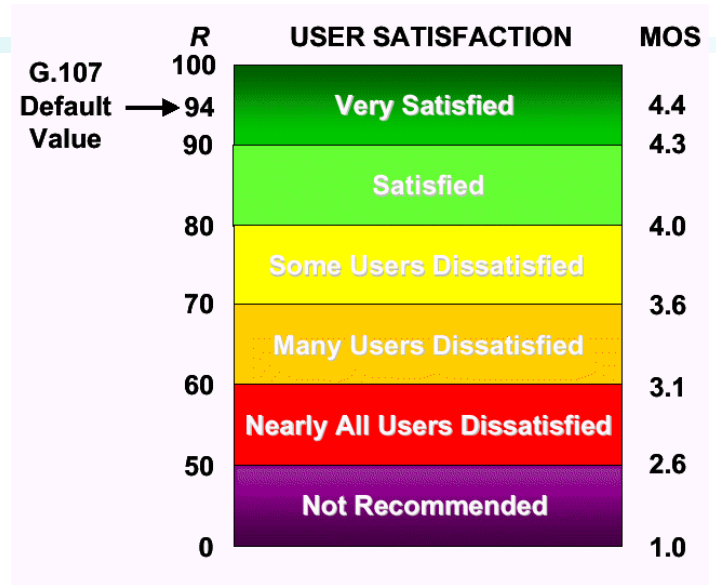
Measured mouth-to-ear delay

VoLTE实测





MOS分值与R因子关系



(From R to MOS, from G.107)

$MOS = 1$	for	$R \leq 0$
$MOS = 1 + 0.035R + R(R - 60)(100 - R)7 \times 10^{-6}$	for	$0 < R < 100$
$MOS = 4.5$	for	$R \geq 100$

(From MOS to R, by 3rd order polynomial)

$$R = 3.026MOS^3 - 25.314MOS^2 + 87.060MOS - 57.336$$



E-model和PESQ比较

	E-model	PESQ
Reference voice sample necessary?	No	Yes
Calculation mainly based on ...	Network performance parameters	Degradation of real voice sample during transmission
Suitable for break-in break-out measurements?	No	Yes
Needs / supports signaling KPIs	No	Yes
Covers ...	Conversational quality	Listening quality

MC 小结：话音质量评估

- ◆ 模拟音频的失真
 - 谐波失真、互调失真、噪声等
- ◆ 数字音频的失真
 - 量化失真、编码失真等
- ◆ 电话话音的失真
 - 时延、回声、沿切割等
- ◆ 话音质量评估
 - PESQ：计算MOS分值
 - E-Model：计算R因子（Rating Factor）
 - MOS与R因子的关系



◆ QoS与QoE概述

◆ 话音质量评估

◆ 视频质量评估

□ 图像和视频的失真

- 图像的失真
- 视频的失真

□ 图像和视频的质量评估方法

- 图像和视频质量评估方法概述
- ITU标准中的主观视频质量评价
- 图像和视频质量的客观评价方法

- 基于PSNR的评价

- 基于HVS (human visual system) 生理特征的评价方法

- 基于结构失真的评价方法

□ 网络视频的质量评估



图像和视频的失真

◆ 空间失真

- 空间域上的失真情况
- 例如块效应、模糊效应、振铃效应、阶梯效应（staircase）、色彩流失（color bleeding）、伪轮廓（false contouring）等

◆ 时间失真

- 发生于连续时间轴上的失真
- 例如蚊式噪声（mosquito noise）、闪烁（flickering）、包丢失、抖动（jitter）、贴片式误码等

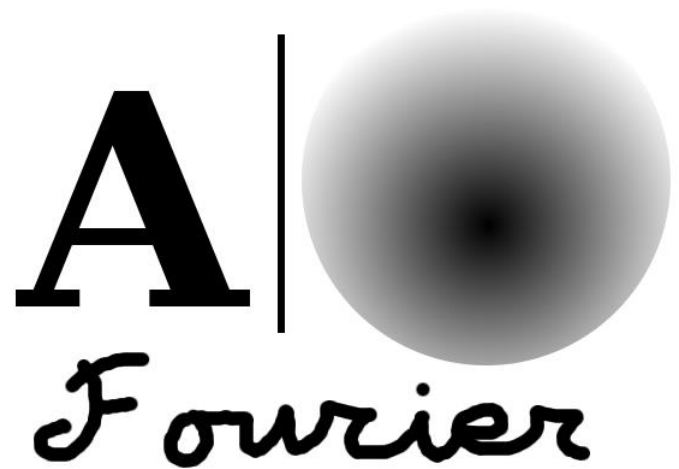
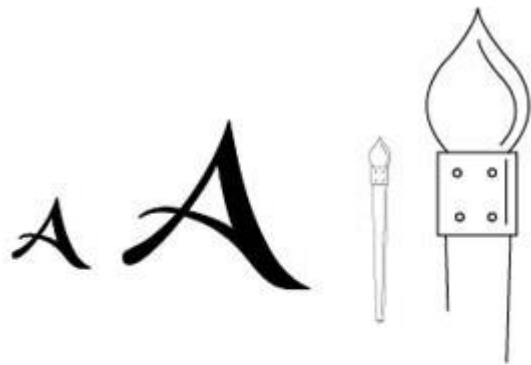
MC 图像的“质量”

◆ 对于“质量”一词并没有统一的定义，确切的定义应该依赖于图像视频类型和应用环境。



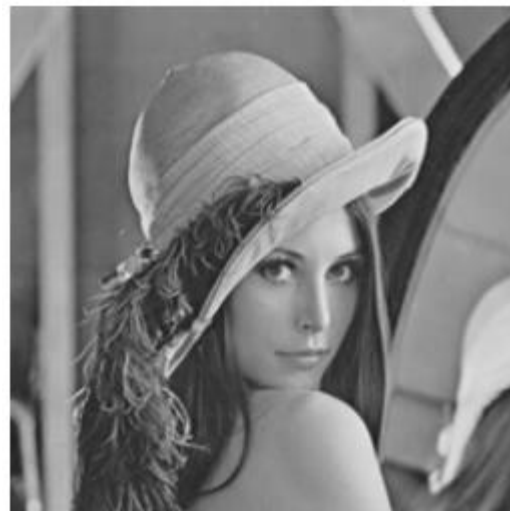
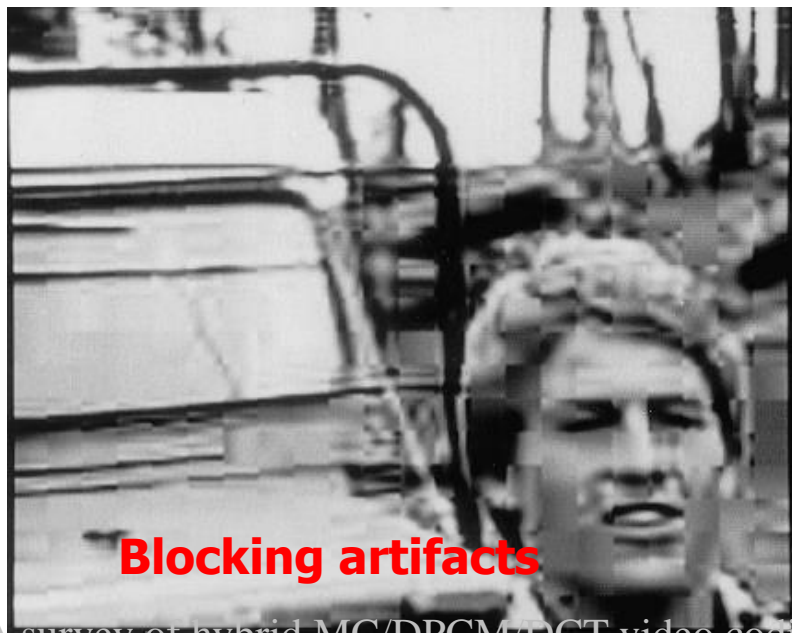
图 1.2-5 图像类型不同，对质量的要求不同

MC Staircase Effect



MC Blocking artifacts

- ◆ The cause of the blocking effect with respect to **block-based coding** is intuitively obvious.



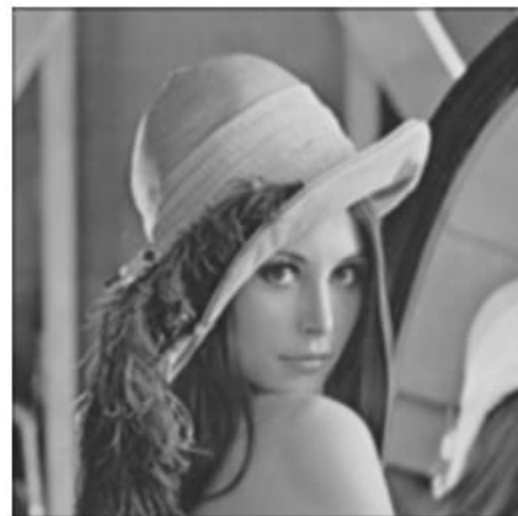
Examples of the blocking effect; most evident in the **smoothly textured regions** which are of **low-to-medium luminance**.

MC Blurring artifacts



Blurring manifests as a loss of spatial detail and a reduction in sharpness of edges in moderate to high spatial activity regions of frames, such as in roughly textured areas or around scene object edges.

Blurring is directly related to the **suppression of the higherorder AC DCT coefficients** through coarse quantization





Color bleeding

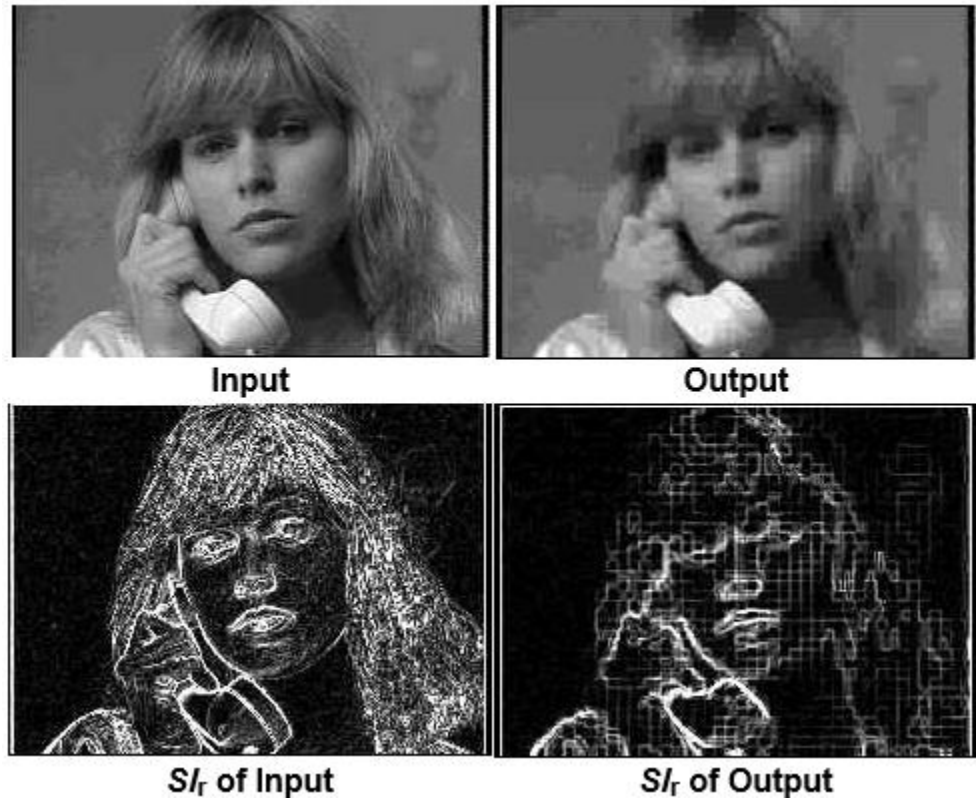
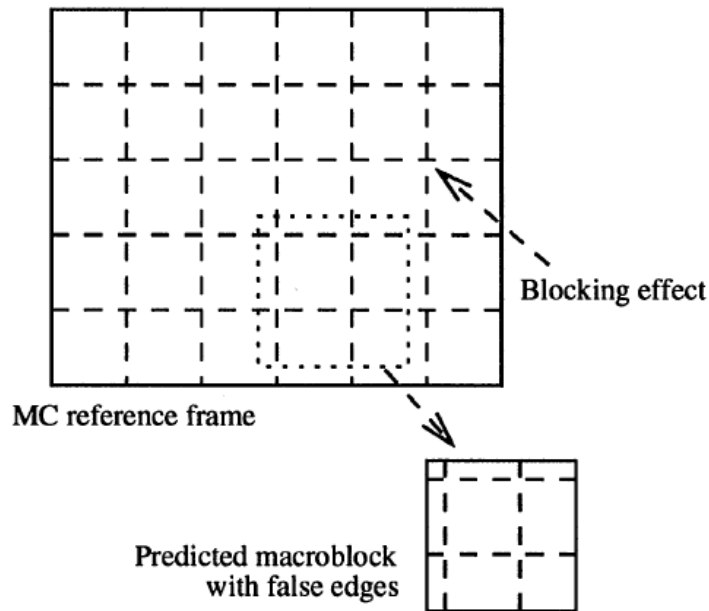
Straight pen-drawn line **color bleeding**, causing jagged edges.



The **blurring of** the luminance information, results in the smoothing of spatial detail. The corresponding effect for the **chrominance** information results in a smearing of the color between areas of strongly contrasting chrominance.

Color bleeding results from the **quantization to zero of the higher-order AC coefficients**, resulting in the representation of the chrominance components with only the lower frequency basis images.

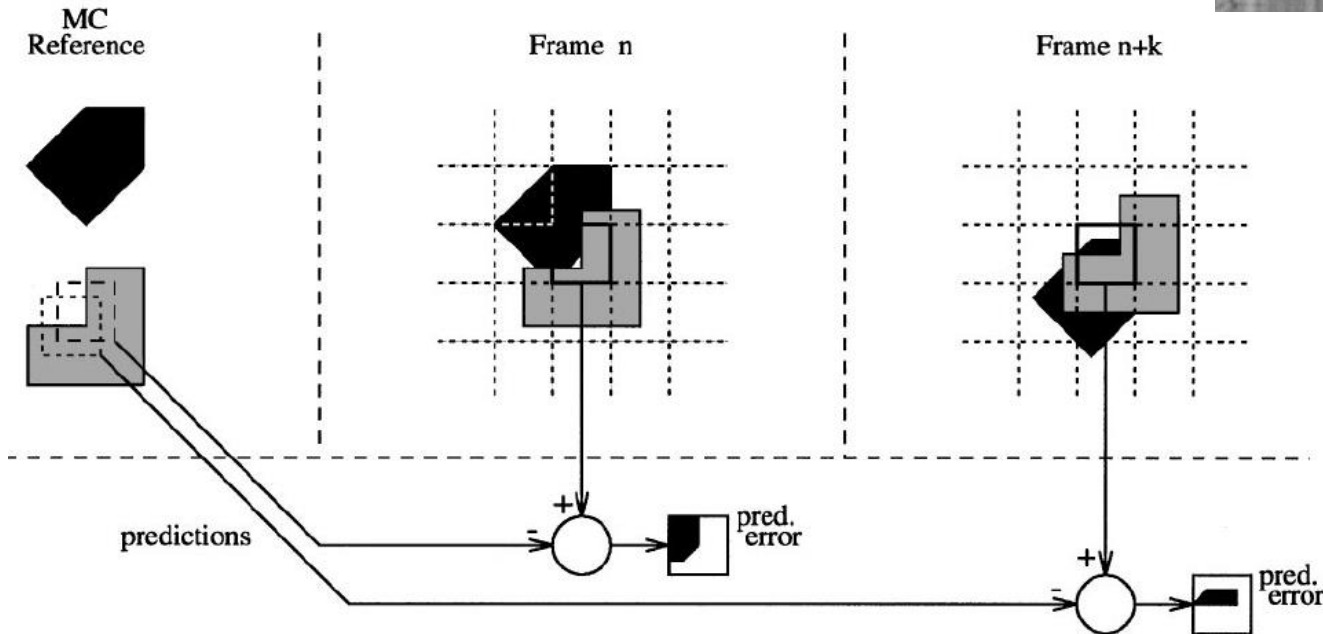
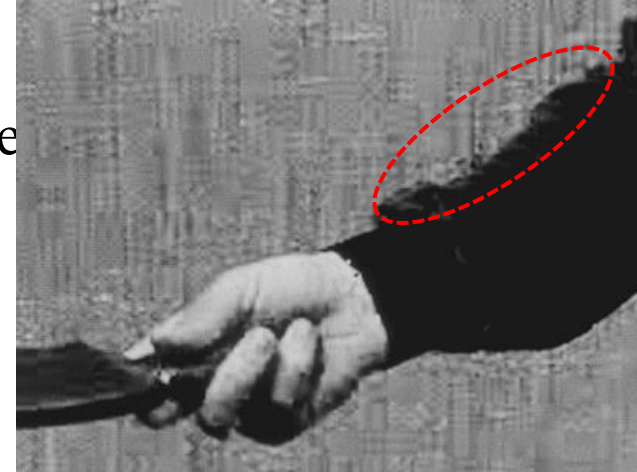
MC False edges



False edges are a consequence of the transfer of the block-edge discontinuities (of MC reference frames) formed by the blocking effect into the current frame.

MC MC mismatch

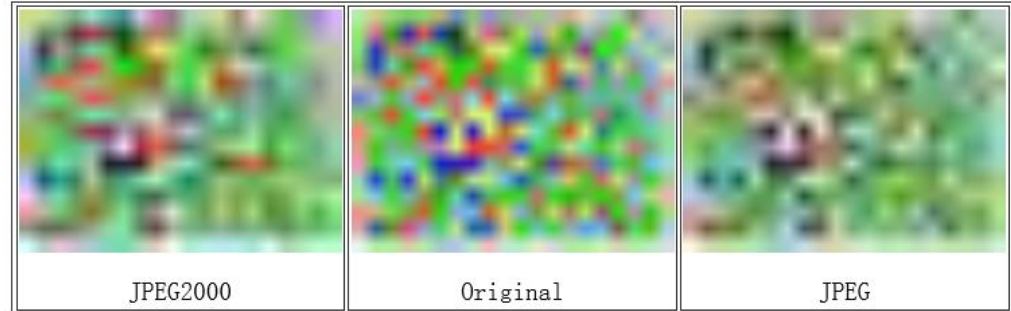
MC mismatch can be generally defined as the situation in which a satisfactory prediction cannot be found for a particular macroblock, resulting in a prediction whose spatial characteristics is mismatched with those of the current macroblock.



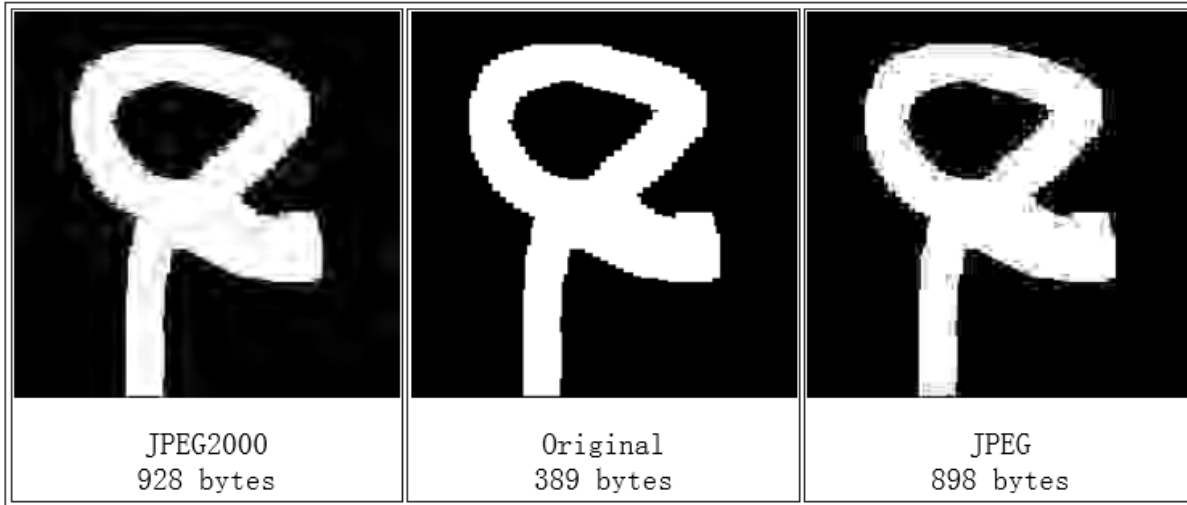


Chrominance mismatch

The use of only the luminance information for the block matching operation results in the selection of the macroblock-sized region within the search window which has the highest luminance correlation to the macroblock currently being coded. This correlation may not extend to the chrominance information, and may even be totally disparate to the chrominance information of the current macroblock.



MC Ringing effect

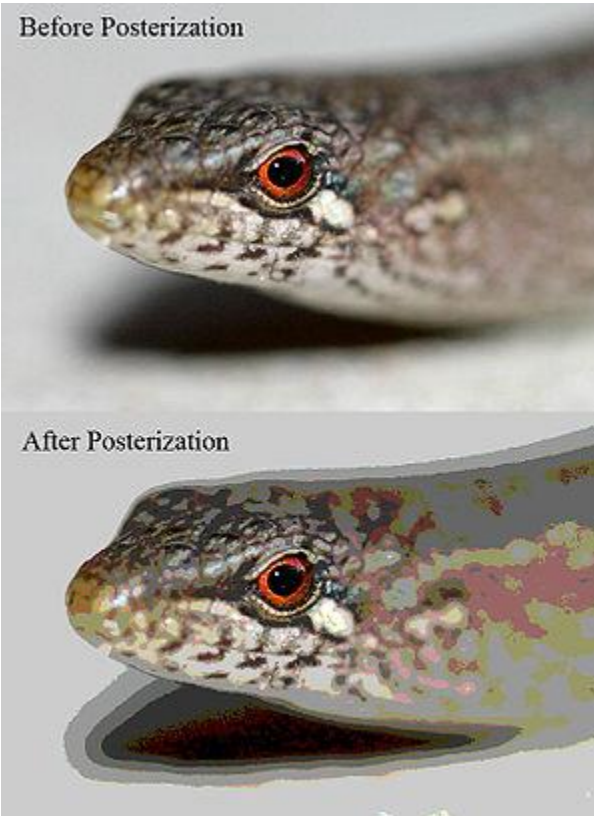


The ringing effect is fundamentally associated with **Gibb's phenomenon**, and, as such, it is most evident along high contrast edges in areas of generally smooth texture in the reconstruction, and appears as a shimmering or rippling outwards from the edge up to the encompassing block's boundary.





False contouring



The artifact is a consequence of the **inadequate quantization** of the **DC coefficient** and the **lower-order AC coefficients** in smoothly textured areas. The effect appears in the reconstruction as step-like gradations in areas of originally smooth transition.



Typical video quality degradation factors

Type of degradation		Degraded appearance	Coding degradation	Transmission degradation
Spatial distortion	Reduced resolution, blurring	Fine patterning, blurred edges, reduced picture resolution	Yes	
	Block distortion	Distortion in the form of a mosaic or geometric pattern	Yes	Yes
	False outline	False outlines appearing in parts where there is a gentle change of color or brightness	Yes	
Temporal distortion	Jerkiness	Smoothness of movement is lost and the picture appears jerky	Yes	Yes
	Flicker	Fluctuations in the intensity level cause the picture to flicker	Yes	
	Motion blur	Moving regions appear blurred	Yes	
	Interruptions and frozen images	Intermittent playback/frozen picture		Yes
Spatio-temporal distortion	Mosquito noise	Mosquito-like buzzing can be heard when the picture contains abrupt edges or changes of color	Yes	
	“Busy” edges	Edges look rough and flickery	Yes	
	Disturbance (failure)	Original image is irrecoverably lost over some or all of the screen		Yes

MC Mosquito noise

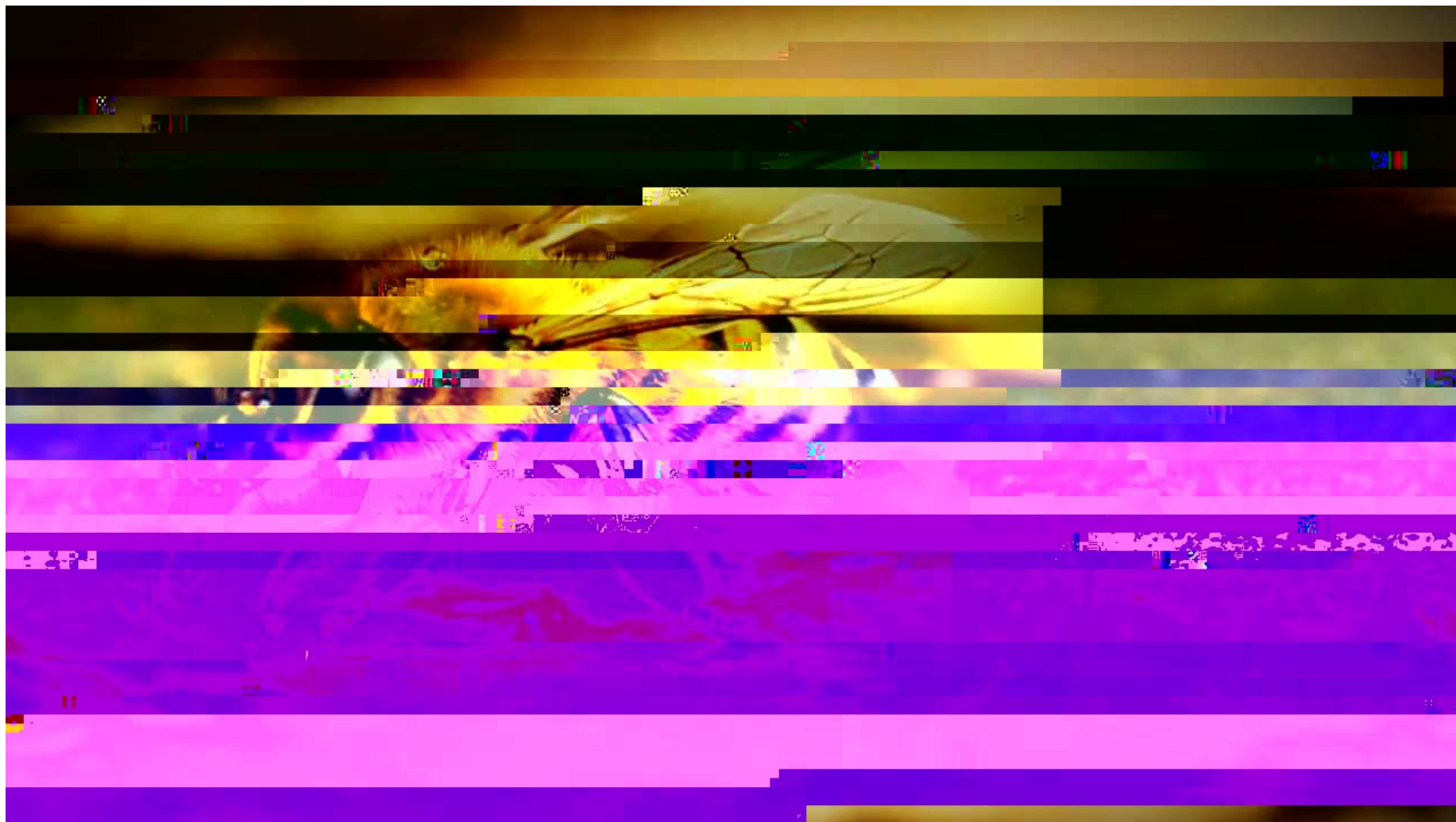
MOSQUITO NOISE



Mosquito noise is produced in image regions with **sharp edges**, such as in the lettering.

The mosaic pattern, typically, coincides with the blocking effect. Coarse quantization will result in the truncation of a significant proportion of the higher-frequency AC coefficients to zero. Consequently, upon reconstruction, the blocks will contain textures constructed only from the lower-frequency AC basis images, which may be of a dissimilar texture or contour than their neighbors.

MC 视频的失真示例



MC 视频的失真示例





(a)



(b)

Figure 1 Examples of compression artifacts. Blockiness resulting from MPEG-2 compression is evident in (a), blur/loss of detail is the main distortion in (b).



编码方式对质量的影响

Original



MPEG-2



H.264





(a)



(b)

Figure 2 Examples of transmission error effects:
Lost blocks and slices in an MPEG-2 bitstream (a), smearing in an M-JPEG2000 clip (b).



各因素对视频质量的直观影响



Reference



Blockiness



Bluriness



Noise



Loss



Jitter

- ◆ QoS与QoE概述
- ◆ 话音质量评估
- ◆ 视频质量评估
 - 图像和视频的失真
 - 图像的失真
 - 视频的失真
 - 图像和视频的质量评估方法
 - 图像和视频质量评估方法概述
 - ITU标准中的主观视频质量评价
 - 图像和视频质量的客观评价方法
 - 基于PSNR的评价
 - 基于HVS (human visual system) 生理特征的评价方法
 - 基于结构失真的评价方法
 - 网络视频的质量评估

MC 视频质量评估方法

◆ 主观评价：MOS

- MOS_V (Mean Opinion Score for Video) 是将图像质量划分为0~5的等级来评定图像质量的好坏，是以用户感知的角度来评价视频质量。

◆ 客观评价

- 基于PSNR和MSE
- 基于HVS
- 基于结构相似度



图像客观质量评价的分类

FR、NR、RR

◆ 全参考 FR（Full Reference）模型

- 算法需要同时输入**参考图像**和失真图像，对输入的图像进行比较，完成质量评价
- 目前**90%以上都是FR模型**

◆ 部分参考 RR（Reduced Reference）模型

- 算法只需要得到参考图像的部分**统计特征**，完成失真图像的质量评价

◆ 无参考 NR（No Reference）模型

- 不需要输入参考图像，根据**失真图像的特征信息**进行质量评价



视频客观质量评价的分类

FR、NR、RR

◆ 视频质量评价模型可以依据上述三种方法进行类似的分类，但更侧重于根据降质原因进行分类，事实上这种方式应该属于特定型模型的进一步细分。

◆ 视频的降质一般来源于编码（coding）和传输（transmitting）两大类因素，或是两者的结合。

◆ 目前已知的视频质量评价模型大多是从图像质量评价模型引申而来，即将视频分解为帧图像，用已有的图像质量评价模型对帧图像进行质量预测，然后综合所有帧质量得到最终的视频质量。

□ 最简单的结合方式是线性平均，但这种方法简单地将每帧独立考虑，类似于在图像中将每一像素独立考虑，在忽视了帧与帧之间的相互关系的情况下这类评价模型并不足以真实地反映出视频的质量，因而所得到的预测结果与人的主观质量感受极有可能不一致。

- ◆ QoS与QoE概述
- ◆ 话音质量评估
- ◆ 视频质量评估
 - 图像和视频的失真
 - 图像的失真
 - 视频的失真
 - 图像和视频的质量评估方法
 - 图像和视频质量评估方法概述
 - ITU标准中的主观视频质量评价
 - 图像和视频质量的客观评价方法
 - 基于PSNR的评价
 - 基于HVS (human visual system) 生理特征的评价方法
 - 基于结构失真的评价方法
 - 网络视频的质量评估



主观视频质量评价的ITU标准

◆主观评价分两大类。第一类评价是确定在最佳条件下系统的性能。这类评价通常称为**质量评价**。第二类评价是确定在与传输或发射有关的非最佳条件下系统维持一定质量的能力。这类评价通常称为**损伤评价**。

(Table 1.3.1) Principal ITU-T/R recommendations relating to subjective video quality assessment methods

Recommendation No.	Title
BT.500	Methodology for the subjective assessment of the quality of television pictures
P.910	Subjective video quality assessment methods for multimedia applications
J.140	Subjective picture quality assessment for digital cable television systems
BT.710 HDTV	Subjective assessment methods for image quality in high-definition television
BT.1129 SDTV	Subjective assessment of standard definition digital television (SDTV) systems
BT.802/BT.1210	Test materials to be used in subjective assessment



ITU-R BT.500-12 建议书(09/2009): 电视图像质量的主观评价方法

◆ 双激励（EBU）法

- 双激励法是一种交替方法，因为在这种方法中，评价者首先看到无损伤的基准图像，然后又看到受损伤的同一图像。随后要求评价者根据第一幅图像来评价第二幅。在持续半小时以内的测试阶段里，向评价者以随机的顺序演示一系列带有随机损伤的图像或序列，涵盖所有必要组合。背景效应无损伤的图像包含在这些待评图像或序列中。在一系列测试阶段结束时，计算每一测试条件和测试图像的平均评分。

◆ 单激励（SS）法

- 在单激励法中，显示单一的图像或一个图像序列，并为评价者提供一份整个演示的索引。测试素材可以只包含测试序列，也可以既包含测试序列，又包含其相应的基准序列。对于后一种情况，基准序列作为一个单独的激励显示，并像其他测试激励那样进行评分。



ITU-T P.910关于视频种类的定义

Table D.1 – Definitions of video classes

TV 0	Loss-less: [ITU-R BT.601], 8-bit per pixel, video used for applications without compression.
TV 1	Used for complete post production, many edits and processing layers, intra-plant transmission. Also used for remote site to plant transmission. Perceptually transparent when compared to TV 0.
TV 2	Used for simple modifications, few edits, character/logo overlays, program insertion, and inter-facility transmission. A broadcast example would be network-to-affiliate transmission. Other examples are a cable system regional downlink to a local head-end and a high quality videoconferencing system. Nearly perceptually transparent when compared to TV 0.
TV 3	Used for delivery to home/consumer (no changes). Other examples are a cable system from the local head-end to a home and medium to high quality videoconferencing. Low artifacts are present when compared to TV 2.
MM 4	All frames encoded. Low artifacts relative to TV 3. Medium quality videoconferencing. Usually ≥ 30 fps.
MM 5	Frames may be dropped at encoder. Perceivable artifacts possible, but quality level useful for designed tasks, e.g., low quality videoconferencing.
MM 6	Series of stills. Not Intended to provide full motion (Examples: Surveillance, Graphics).



电视图像主观评价的打分表

◆ Examples of rating dimensions which may be assumed to define factors that contribute to the perceived global image quality are listed below, together with an indication of whether a factor contributes positively or negatively to quality:

- – Brightness (positive);
- – Contrast (positive);
- – Colour reproduction (positive);
- – Outline definition (positive);
- – Background stability (positive);
- – Speed in image reassembling (positive);
- – Jerkiness (negative);
- – "Smearing" effects (negative);
- – "Mosquito" effects (negative);
- – Double images/shadows (negative);
- – Halo (negative).

◆ Recent research has shown that these factors may be combined into a predicted global quality by giving appropriate weightings to each factor and then adding them together.

◆ QoS与QoE概述

◆ 话音质量评估

◆ 视频质量评估

□ 图像和视频的失真

- 图像的失真
- 视频的失真

□ 图像和视频的质量评估方法

- 图像和视频质量评估方法概述
- ITU标准中的主观视频质量评价
- 图像和视频质量的客观评价方法

- 基于PSNR的评价

- 基于HVS (human visual system) 生理特征的评价方法

- 基于结构失真的评价方法

□ 网络视频的质量评估



客观视频质量评价的ITU标准

- ◆ ITU-T J.140 (1998), Subjective picture quality assessment for digital cable television systems.
- ◆ ITU-T J.143 (2000), User requirements for objective perceptual video quality measurements in digital cable television.
- ◆ ITU-T J.144 (2001) “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.”
- ◆ ITU-T J.149 (2004) “Method for specifying accuracy and cross-calibration of video quality metrics (VQM).”
- ◆ ITU-T P.910 (1996), Subjective video quality assessment methods for multimedia applications.
- ◆ ITU-R BT.1683 (2004) “Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference.”

ITU-T J.144, Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference
<http://www.itu.int/rec/T-REC-J.144-200103-S/en>

Objective perceptual assessment of video quality: Full reference television

http://www.itu.int/ITU-T/studygroups/com09/docs/tutorial_opavc.pdf, Retrieved: 2013.11



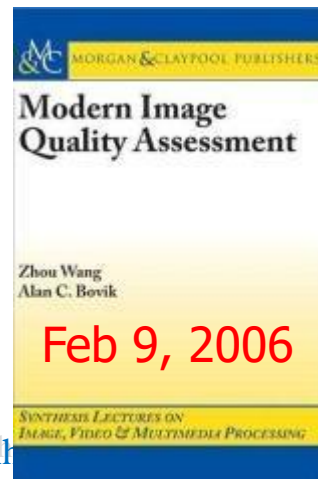
图像视频质量评价领域活跃的实验室

◆ 国际

- 美国德州大学奥斯汀分校的LIVE实验室
- <http://live.ece.utexas.edu/>
- 该实验室的成员Alan Bovik、Zhou Wang、Hamid Rahim Sheikh、Kalpana Seshadrinathan 及 Anush Krishna Moorthy 等均在图像质量评价领域有着出色的工作，例如 Zhou Wang 提出的在图像质量评价方面起着重要推动作用的 **SSIM** 方法，该实验室另一值得称赞的贡献是提供了比较完备的图像主观测试数据库以及视频主观数据库。
-

◆ 国内

- 西安电子科技大学智能感知与图像理解实验室
- <http://see.xidian.edu.cn/iiip/>
-





图像视频客观质量评价方法的趋势 2016

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

Application-driven **no reference quality assessment for dermoscopy images with multiple distortions**

F. Xie, Y. Lu, A.C. Bovik, Z. Jiang and R. Meng

IEEE Transactions on Biomedical Engineering , vol. 63, no. 6, pp. 1248-1256, June, 2016.

Temporal Video Quality Model Accounting for Variable Frame Delay Distortions

M. H. Pinson, L. K. Choi, and A. C. Bovik

IEEE Transactions of Broadcasting , vol. 60, no. 4, pp. 637-649, December, 2014.

A model of the time-varying subjective quality of **HTTP video streams**

C. Chen, L.K. Choi, G. de Veciana, C. Caramanis, R.W. Heath, Jr. and A.C. Bovik

IEEE Transactions on Image Processing, January 2014

Blind prediction of natural video quality

M. Saad and A.C. Bovik

IEEE Transactions on Image Processing, December 2013

Video Quality Assessment by Reduced Reference **Spatio-temporal Entropic Differencing**

R. Soundararajan and A. C. Bovik

IEEE Transactions on Circuits and Systems for Video Technology , vol. 23, no. 4, pp. 684-694, April 2013

(Winner of the IEEE Circuits and Systems for Video Technology **Best Paper Award for 2016**).

Video quality assessment on **mobile devices: Subjective, behavioral, and objective studies**

A. K. Moorthy, L. K. Choi, A. C. Bovik and G. de Veciana

IEEE Journal of Selected Topics in Signal Processing, Vol: 6 No: 6, October 2012 {ynh,cxh}@ustc.edu.cn



图像视频客观质量评价方法的趋势 2017

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

- C. G. Bampis, P. Gupta, R. Soundararajan and A. C. Bovik, "SpEED-QA: **Spatial Efficient Entropic Differencing** for Image and Video Quality," in IEEE Signal Processing Letters, vol. 24, no. 9, pp. 1333-1337, Sept. 2017.
- L. Liu, B. Liu, C. C. Su, H. Huang, A. C. Bovik, "**Binocular spatial activity and reverse saliency** driven no-reference stereopair quality assessment", Signal Processing: Image Communication, 2017.
- C. G. Bampis, Z. Li and A. C. Bovik, "Continuous Prediction of Streaming Video QoE using **Dynamic Networks**," Signal Processing Letters, to appear.
- C. G. Bampis and A. C. Bovik, "**Learning** to Predict Streaming Video QoE: Distortions, Rebuffering and Memory," Transactions on Image Processing, under review.
- D. Kundu, D. Ghadiyaram, A.C. Bovik and B.L. Evans, "No-reference quality assessment of **high dynamic range** pictures," IEEE Transactions on Image Processing, to appear.
- L. Liu, Y. Hua, Q. Zhao, H. Huang and A.C. Bovik, "Blind image quality assessment by relative gradient statistics and Adaboosting **neural network**," Signal Processing: Image Communication , vol. 40, no. 1, pp. 1-15, January, 2016
- D. Ghadiyaram and A. C. Bovik, "Perceptual Quality Prediction on Authentically Distorted Images Using a **Bag of Features** Approach," <http://arxiv.org/abs/1609.04757>

MC 基于PSNR的评价

◆采用峰值信噪比(peak signal noise ratio , PSNR) 或均方差(mean square error , MSE) 衡量视频序列的失真度。

$$MSE = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N [x(m, n) - y(m, n)]^2$$

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N [x(m, n) - y(m, n)]^2}$$

◆缺点：PSNR 和MSE 忽略了图像内容对人眼的影响，不能完整地反映出图像的质量

客观指标与主观感知的差异？



峰值信噪比=5.98

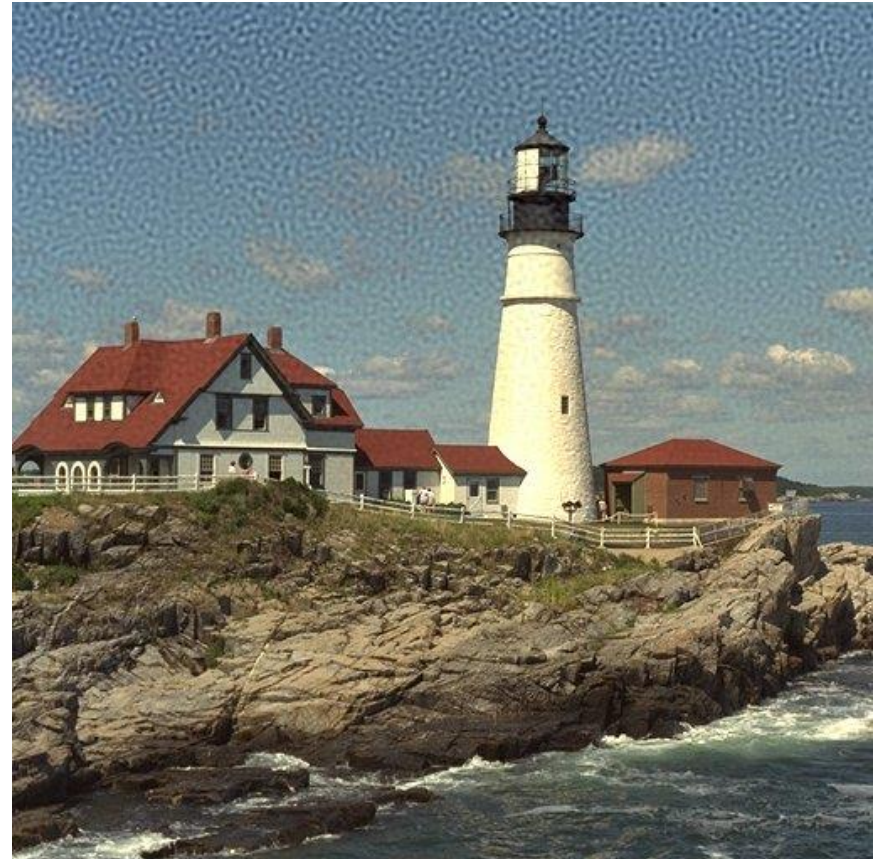
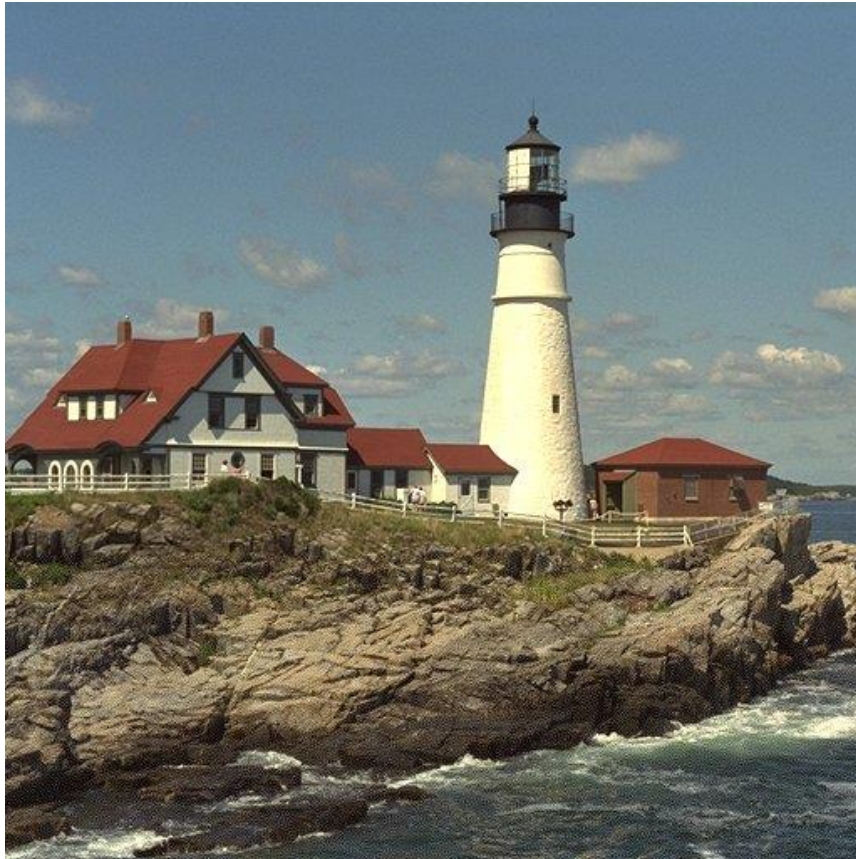


峰值信噪比=6.24

- 结论：
- 1、峰值信噪比度量与人的视觉感知并不完全一致！
 - 2、需要寻求更加符合人类视觉感知的客观度量方法以及相应的编码理论和方法。

MC PSNR vs. QoE

Same amount of distortion (PSNR) – different perceived quality



Understand & model human vision system

Source <http://qoe.symmetricom.com> Sep 2008

{ynh, cxh}@ustc.edu.cn

MC PSNR vs. QoE



PSNR数值都是27.123

MC 改进1: 真实的最大最小值

◆实际图像中,并不一定(或很难)都存在最亮点或最暗点。在这种情况下应以输入图像实际存在的亮度(或色度)最大值和最小值进行计算才是正确的。

$$10 \log [Y_{o, \max} (k) - Y_{o, \min} (k)]^2 / M_{se} \quad (1)$$

$$M_{se} = \frac{1}{N} \sum_i \sum_j [y_o (i, j, k) - y_i (i, j, k)]^2 \quad (2)$$

◆其中,N 为输入图像或输出图像第k 帧的像 $Y_{o, \max} (k)$ 、 $Y_{o, \min} (k)$ 分别为输出图像第k 帧中最大、最小像素亮度的值



改进2：斜空间滤波求模糊系数

◆即分别对原始图像和重建图像用微分型空间滤波器提取其高频成份，再求二者的相对差值，定义为“模糊系数”

◆选斜方向的微分型空间滤波器

$$K_{\text{blur}} = S_{i,\text{out}} / S_{i,\text{in}}$$

$$S_i = \sum_j \sum_k |y_f'(i, j, k)|$$

-1	0	1
0	0	0
1	0	-1

◆ K_{blur} 值的范围一般在0~1 之间

◆k 值越接近于1，图像清晰度越高



改进3：正空间滤波求方块效应

◆对分块进行压缩的编码，块效应(Mosaic)的出现增加了重建图像中的垂直和水平边缘, 如果单纯用微分型垂直和水平二维空间滤波器滤出垂直与水平方向的边缘, 则必与图像内容中真正存在的垂直与水平边缘混在一起

◆选用正空间滤波器

$$K_{\text{block}} = 1 - K_{\text{blur}} S_{V, \text{in}} / S_{V, \text{out}}$$

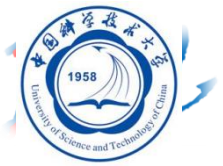
$$S_V = \sum \sum |y_e'(i, j, k)|$$

0	1	0
-1	0	-1
0	1	0

◆ K_{block} 系数的大小一般在0~1之间, 方块效应越明显, 该系数就越接近于1

MC 基于HVS 的评价方法

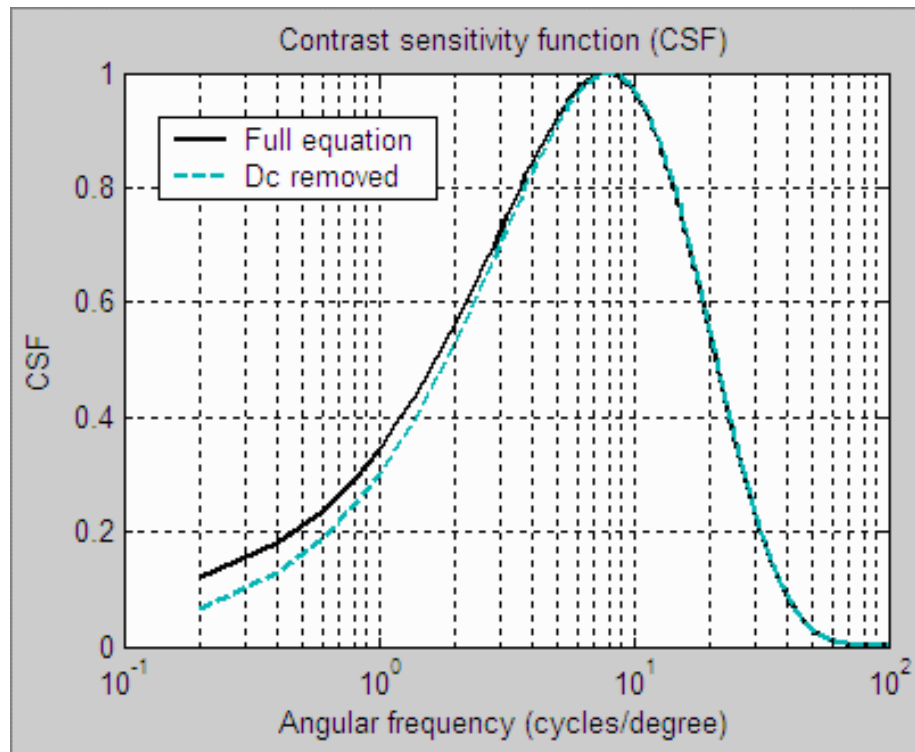
- ◆ HVS (human visual system)
- ◆ 图像最终接受者是人，所以好的图像质量评价方法应考虑人的视觉屏蔽特性。
 - 应用视觉灵敏度函数(CSF)
 - MPQM(Moving Pictures Quality Metric)
 -
- ◆ 基于视觉感兴趣区的视频质量评价方法



对比度敏感函数

Contrast Sensitivity Function (CSF)

在空间或时间上两相邻的视觉信号，人眼刚刚能鉴别出两者差别的能力称为**视觉系统的分辨率**。分辨率可用**CSF**来表示。



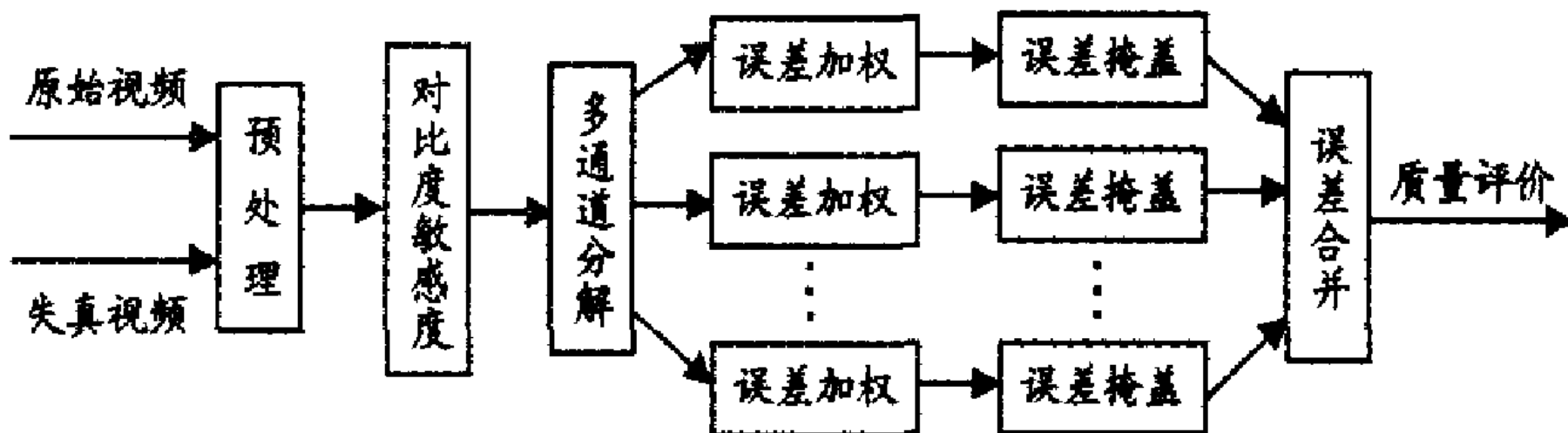
$$\text{CSF}(f) = 2.6 (0.0192 + 0.114 f) \exp(-0.114 f)^{1.1}$$

The dc term (0.0192) can be dropped with very little effect.

$$\text{CSF}(f) = (0.0192 + 0.114 f) \exp(-0.1254 f)$$

MC HVS方法原理框图

◆原始视频和失真视频都要经过空间校准、色彩空间转换、显示校准、亮度调节等**预处理**，由对比度和灵敏度函数(CSF)对引起对比度变化的颜色、时间、空间等激励因素进行带通滤波，再**分解**为多个具有带通滤波特性的空间频率**通道**和时间频率通道。视频机制的多通道之间存在着相互作用，但目前多通道之间的相互作用机制尚不明确。





Moving Pictures Quality Metric (MPQM)

- ◆ seventeen spatial filters and two temporal filters. (34 perceptual components)
- ◆ using a contrast sensitivity function estimation and a linear summation model of masking
- ◆ computation of the distortion E, Where $e(x,y,t,c)$ is the masked error signal at position (x,y) and time t in the current block and in the channel c ; N_x, N_y, N_t are the horizontal and vertical dimensions of the blocks; N is the number of channels. β is a constant having the value 4.

$$E = \left(\frac{1}{N} \sum_{c=1}^N \left(\frac{1}{N_x N_y N_t} \sum_{t=1}^{N_t} \sum_{y=1}^{N_y} \sum_{x=1}^{N_x} |e(x, y, y, c)| \right)^\beta \right)^{\frac{1}{\beta}},$$

- ◆ Masked PSNR (MPSNR)

$$MPSNR = 10 \log \frac{255^2}{E^2},$$

$$Q = \frac{5}{1 + \gamma E}$$



基于结构失真的评价方法

Structural SIMilarity (SSIM) Index

- ◆ 结构相似(structural similarity , SSIM) 法
- ◆ 认为人眼视觉的主要功能是提取场景中的结构信息, 提出以结构相似度来评价图像质量。
- ◆ SSIM利用高层视觉对结构敏感的特性, 从图像局部亮度、对比度、结构三个层次的相似程度, 描述了原始图像与失真图像的相似性。SSIM模拟“自上而下”的HVS模型, 从高层视觉出发理解图像质量, 而回避了直接建立复杂的底层视觉模型, 包括HVS模型参数、超阈问题、失真与自然图像的复杂性等等都不再是评价图像质量的障碍。可以说, SSIM以一种简洁美妙的方式较好地评价了图像质量。

<http://www.cns.nyu.edu/~lcv/ssim/> 提供matlab代码

[PDF]Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 截止2014年11月被引用次数: 8363

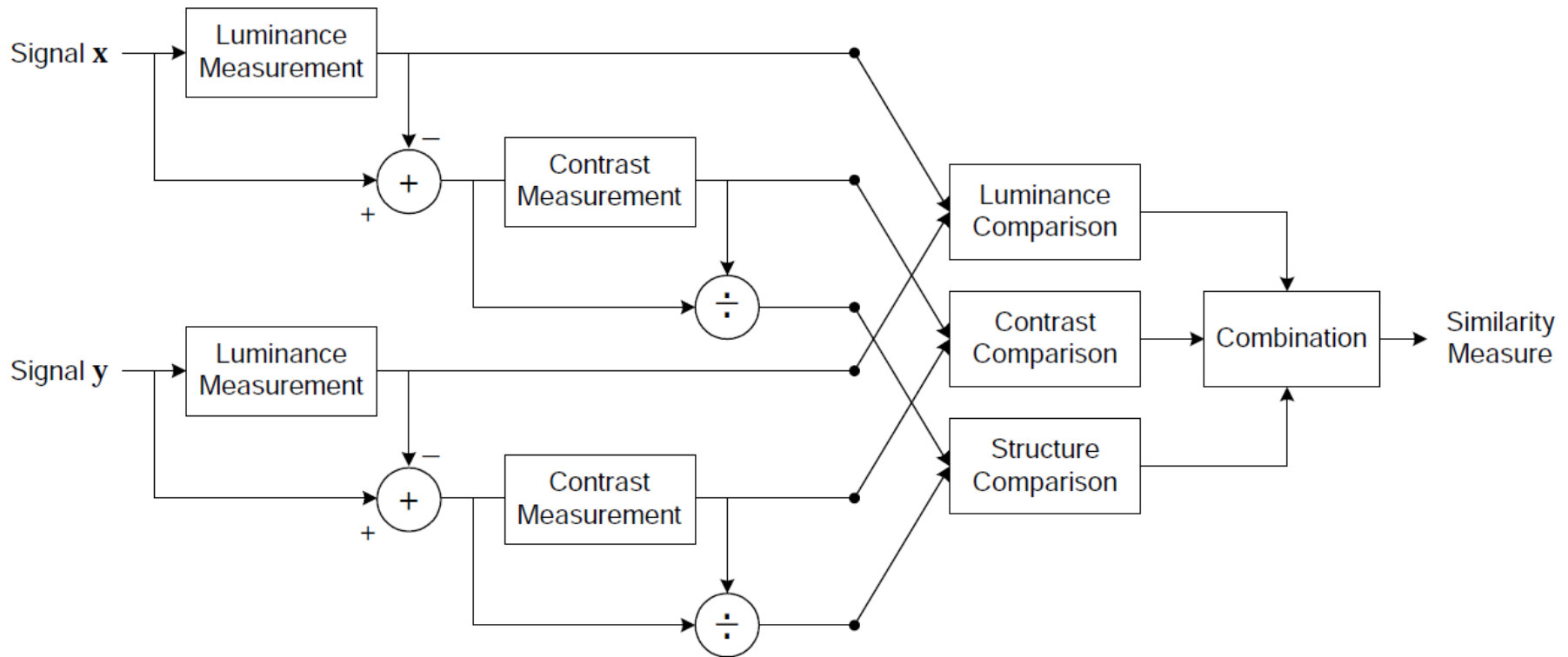
103

{ynh, cxh}@ustc.edu.cn

IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004



基于SSIM的相似度测量



图像之间的质量相似度被分成为三个部分：亮度相似度、对比度相似度和结构相似度。



SSIM中相似度的定义

◆在SSIM算法中，图像之间的质量相似度被分成三个部分：亮度相似度、对比度相似度和结构相似度。

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

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

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

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

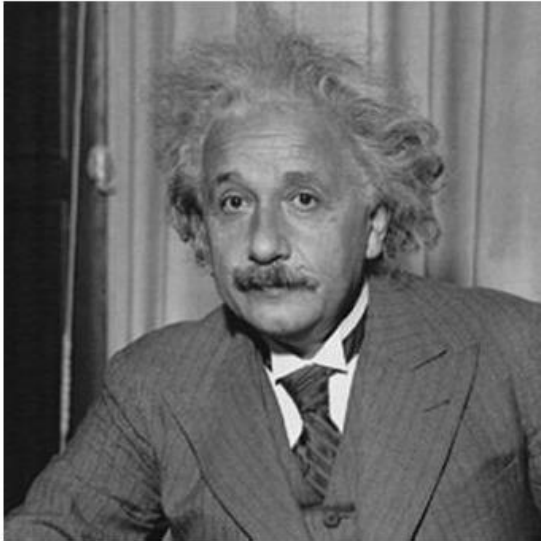
$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

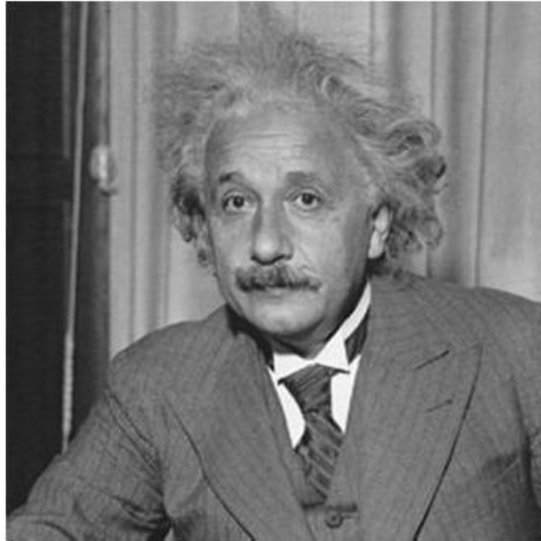
$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$



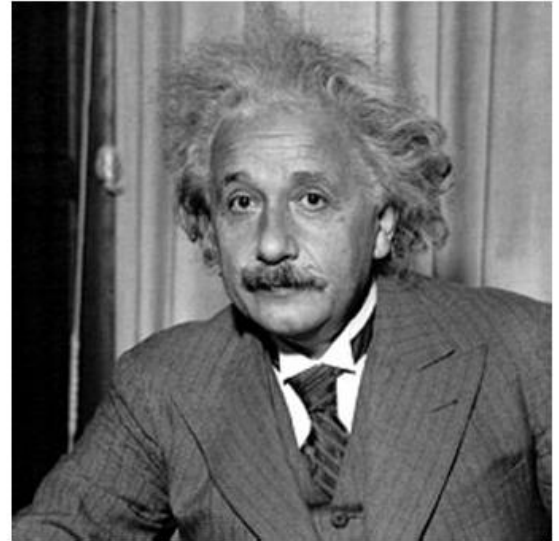
SSIM的有效性 vs. PSNR



Original, MSE = 0; SSIM = 1



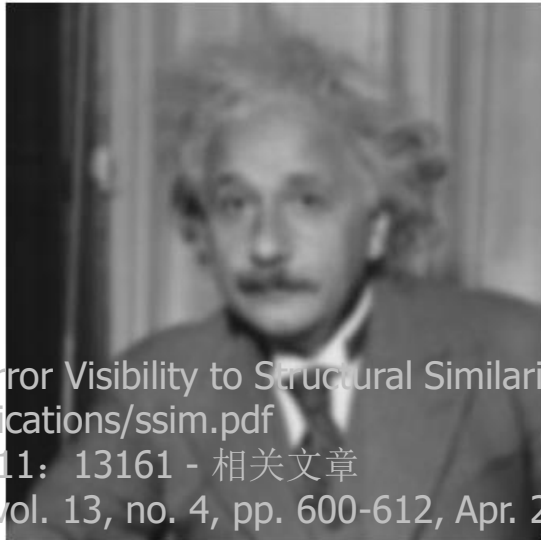
MSE = 144, SSIM = 0.988



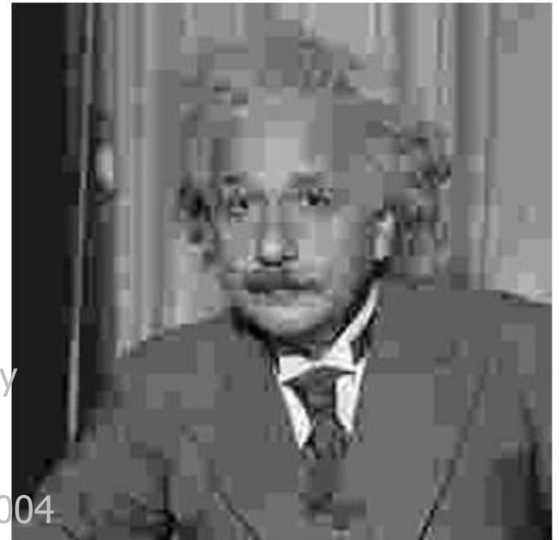
MSE = 144, SSIM = 0.913



MSE = 144, SSIM = 0.840



MSE = 144, SSIM = 0.694



MSE = 142, SSIM = 0.662

<http://www.cns.nyu.edu/~lcv/ssim/>

[PDF] Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 被引用次数2016.11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004



SSIM的有效性



(a)



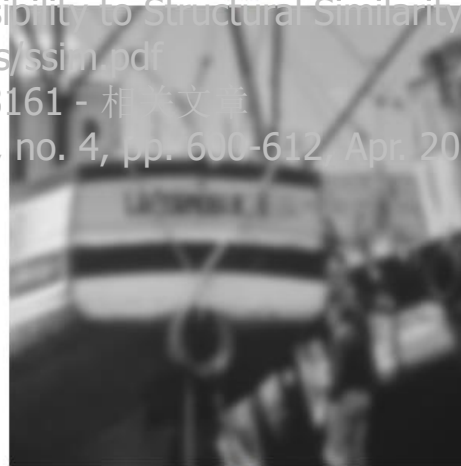
(b)



(c)



(d)



(e)



(f)

Fig. 2. Comparison of “Boat” images with different types of distortions, all with $MSE = 210$. (a) Original image (8bits/pixel; cropped from 512×512 to 256×256 for visibility); (b) Contrast stretched image, $MSSIM = 0.9168$; (c) Mean-shifted image, $MSSIM = 0.9900$; (d) JPEG compressed image, $MSSIM = 0.6949$; (e) Blurred image, $MSSIM = 0.7052$; (f) Salt-pepper impulsive noise contaminated image, $MSSIM = 0.7748$.

<http://www.cns.nyu.edu/~lcv/ssim/>

[PDF]Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 被引用次数2016-11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004



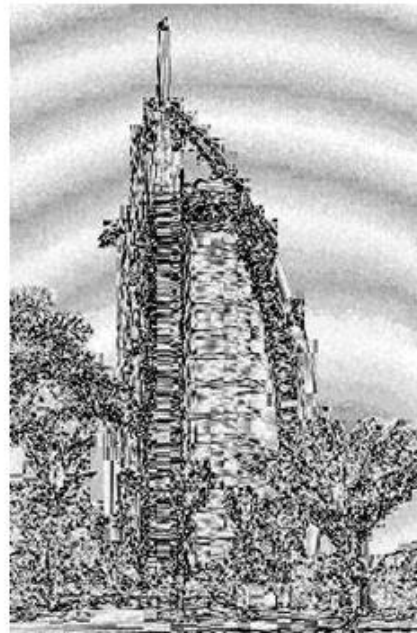
SSIM index map



Original image



JPEG compressed image



Absolute error map



SSIM index map

<http://www.cns.nyu.edu/~lcv/ssim/>

[PDF]Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 被引用次数2016.11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004



不同失真情况下的SSIM



Original image

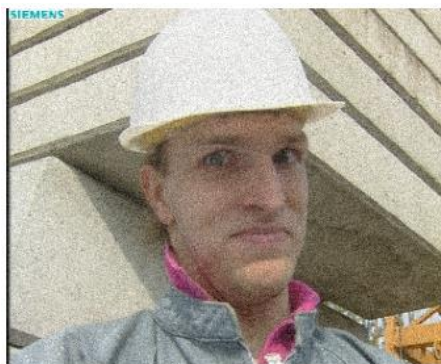


Image with added noise



Blurred image



Sharpen image



SSIM = 1



SSIM = 0.552119



SSIM = 0.9225

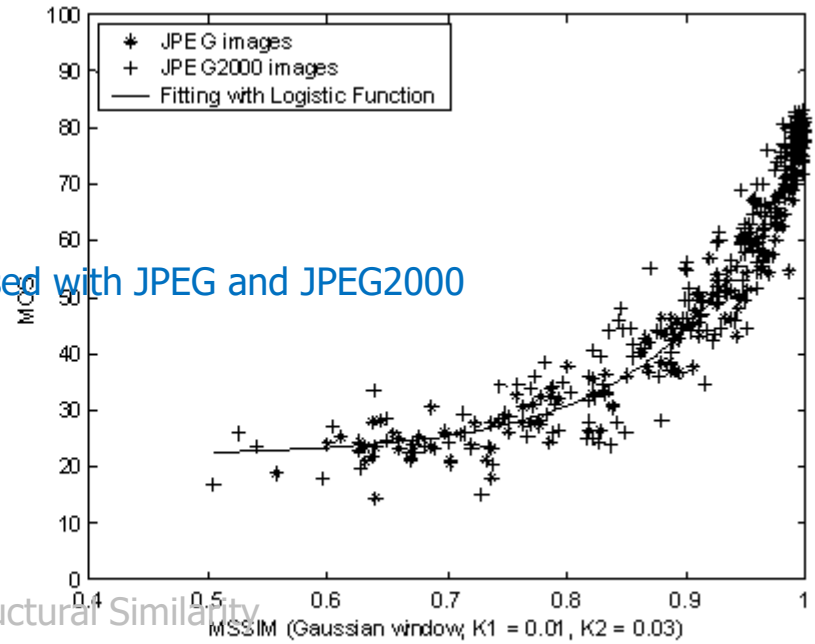
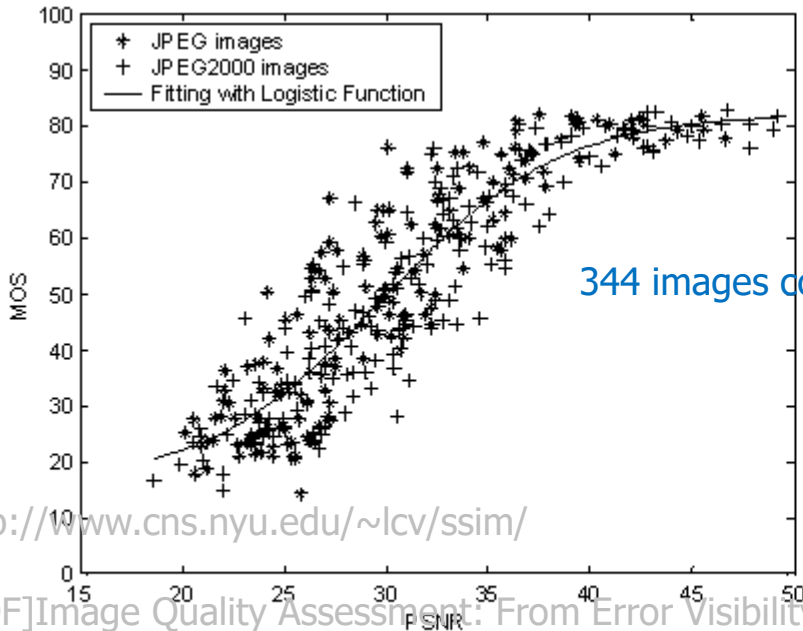


SSIM = 0.958917



SSIM与主观评价的关系

The scatter plots of the subjective measurement (mean opinion score, MOS) versus the objective predictions (PSNR and MSSIM) are shown below, where each point represent one test image. Clearly, **MSSIM is much better** in predicting the perceived image quality.



MSSIM vs. MOS

<http://www.cns.nyu.edu/~lcv/ssim/>

[PDF]Image Quality Assessment: From Error Visibility to Structural Similarity
<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 被引用次数2016.11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp.600-612, Apr. 2004



Structural SIMilarity (SSIM) Index

◆ SSIM 方法形式简单，并且其预测性能相比于以往方法得到了极为显著的提升。结构相似度量方案为图像视频质量评价领域提供了新的研究思路，并且得到了众多关注，这在一定程度上极大地促进了该领域的新发展。

◆ 围绕着 SSIM 的改良和扩展

- 空间域的多尺度 SSIM (MS-SSIM)
- 频率域的复小波 SSIM (CW-SSIM)
- 小波 SSIM (DWT-SSIM)
- 基于结构信息提取的图像质量评价 SIExt (Structural Information Extraction)

□ 结合图像内容与 SSIM 的 CBM

<http://www.cns.nyu.edu/~jvcv/ssim/>

[PDF] Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者: Z Wang - 2004 - 被引用次数 2016.11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004



小结：图像质量的客观评价

- ◆ 核心是相似度的定义（Metric）
 - MSE/PSNR：欧式空间的距离
 - Moving Pictures Quality Metric (MPQM)
 - seventeen spatial filters and two temporal filters
 - SSIM：亮度相似度、对比度相似度和结构相似度

<http://www.cns.nyu.edu/~lcv/ssim/>

[PDF]Image Quality Assessment: From Error Visibility to Structural Similarity

<https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf>

作者：Z Wang - 2004 - 被引用次数2016.11: 13161 - 相关文章

IEEE Transactions on Image Processing, vol. 13, no. 4, pp.600-612, Apr. 2004

- ◆ QoS与QoE概述
- ◆ 话音质量评估
- ◆ 视频质量评估
 - 图像和视频的失真
 - 图像的失真
 - 视频的失真
 - 图像和视频的质量评估方法
 - 图像和视频质量评估方法概述
 - ITU标准中的主观视频质量评价
 - 图像和视频质量的客观评价方法
 - 基于PSNR的评价
 - 基于HVS (human visual system) 生理特征的评价方法
 - 基于结构失真的评价方法
 - 网络视频的质量评估

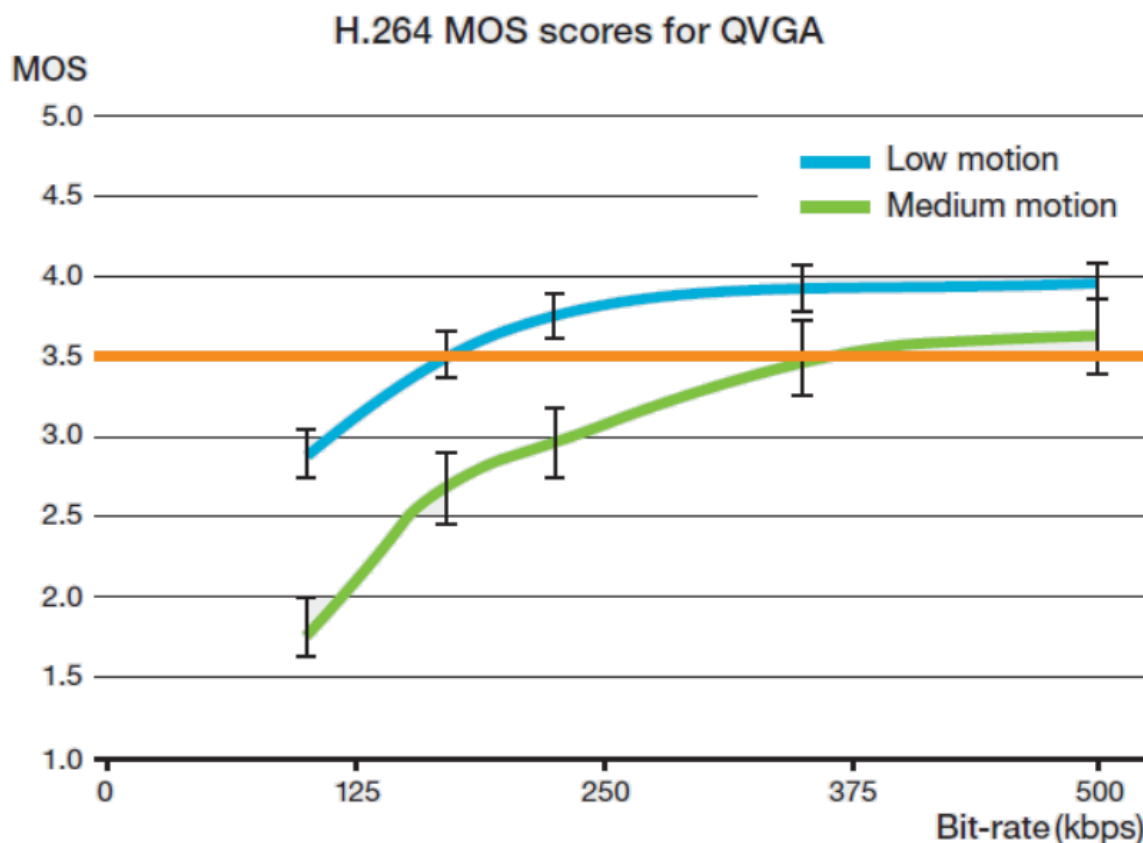


VoLTE网络测试实例

QVGA视频在VoLTE网络中的质量

对于低速和中等速度变化的视频而言，传输过程丢包率和可感知的视频质量并不是线性关系

FIGURE 3 Video quality versus video bit-rate for low and medium motion content



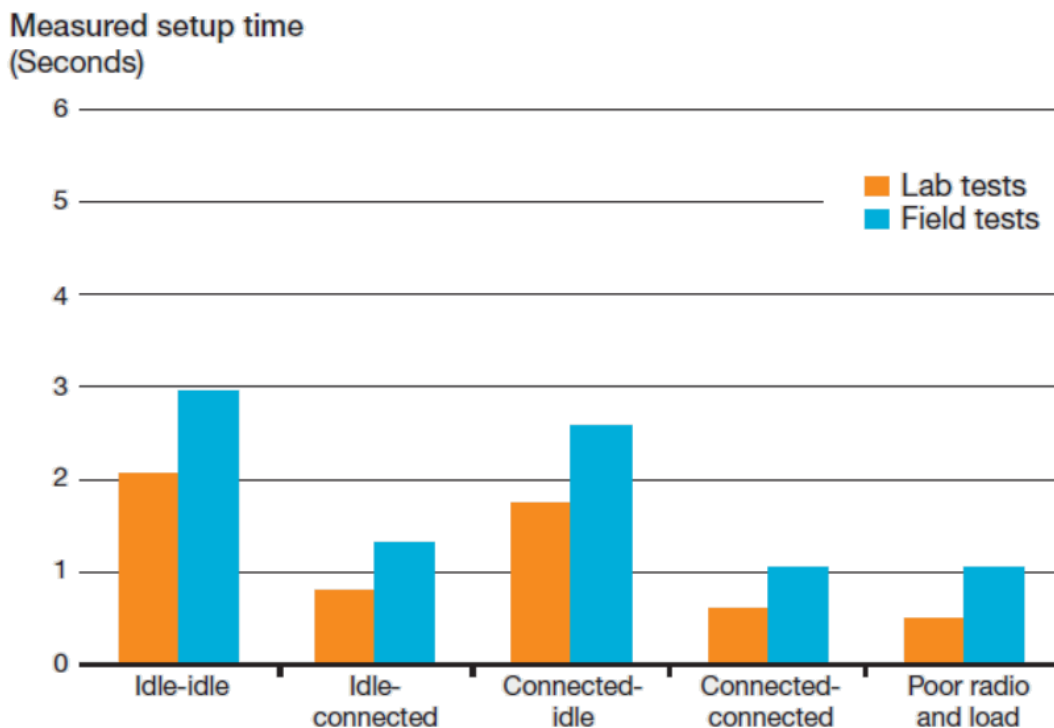


VoLTE网络测试实例

实验与实测的Call setup time

VoLTE的呼叫建立时间测试（call-setup-time）表明，VoLTE的呼叫建立时间明显短于WCDMA。图中实验室测试结果和现场测试结果的差异与终端、核心网的具体部署有关系。

FIGURE 5 Measured call setup time



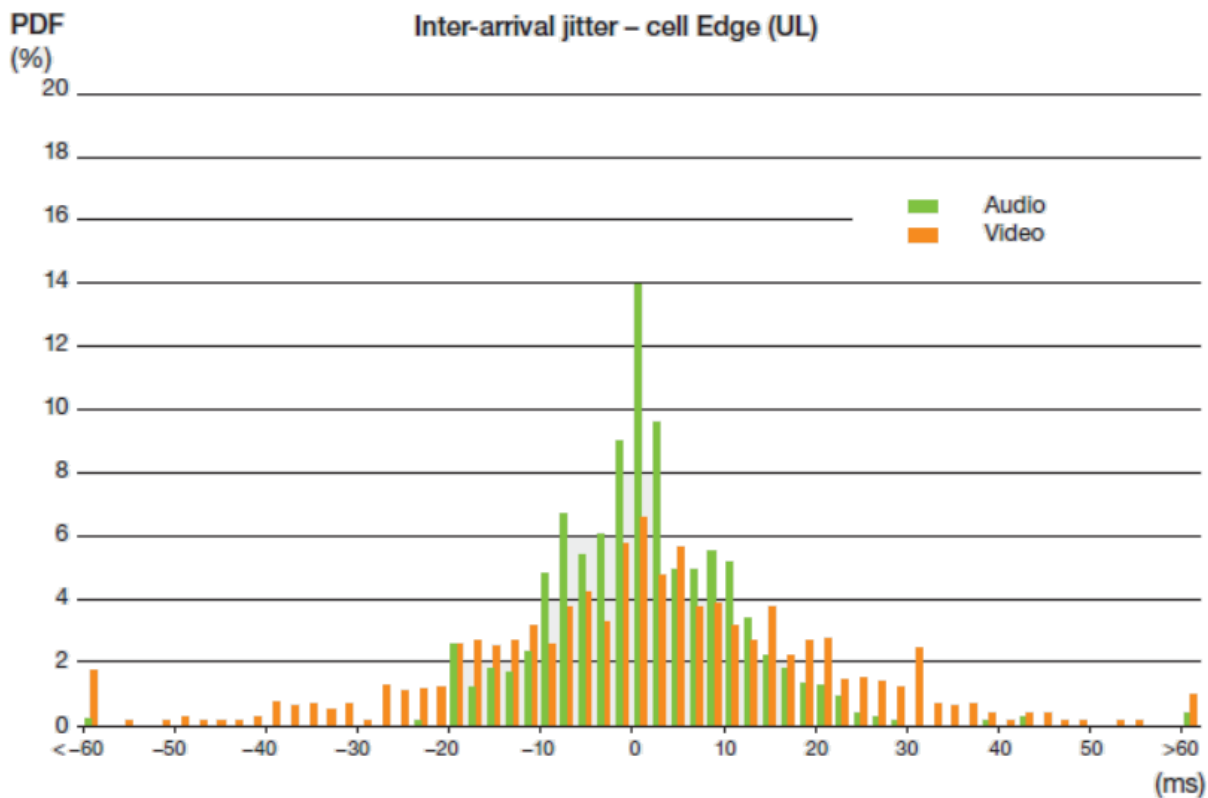


VoLTE网络测试实例

VoLTE网络的实测jitter

VoLTE网络抖动测试表明，视频电话的抖动高于音频电话的抖动。这和视频流中大数据包尺寸有密切的关系。

FIGURE 8 UL inter-arrival jitter measured in the field trial - video telephony





丢包与MPEG2视频质量关系的分析

MPEG2编码序列的数据结构

◆ MacroBlock

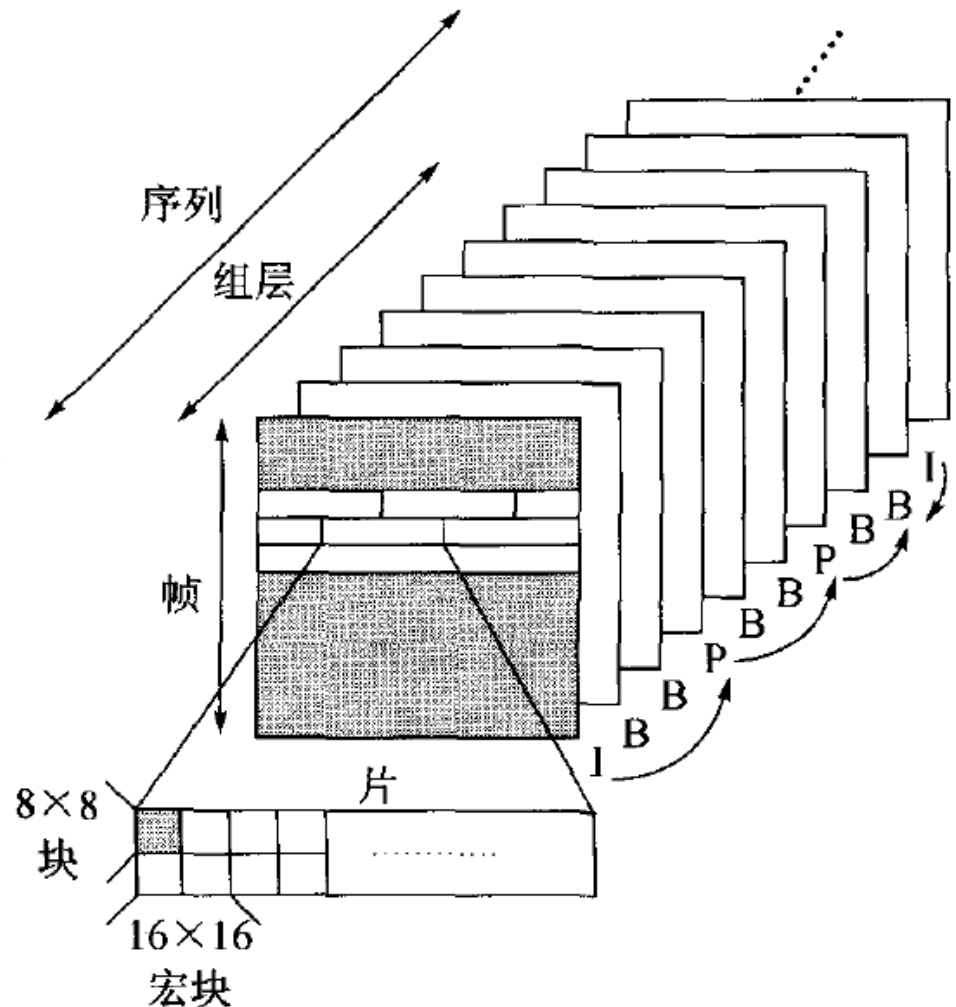
◆ Slice

◆ Frame

□ I Frame

□ P Frame

□ B Frame

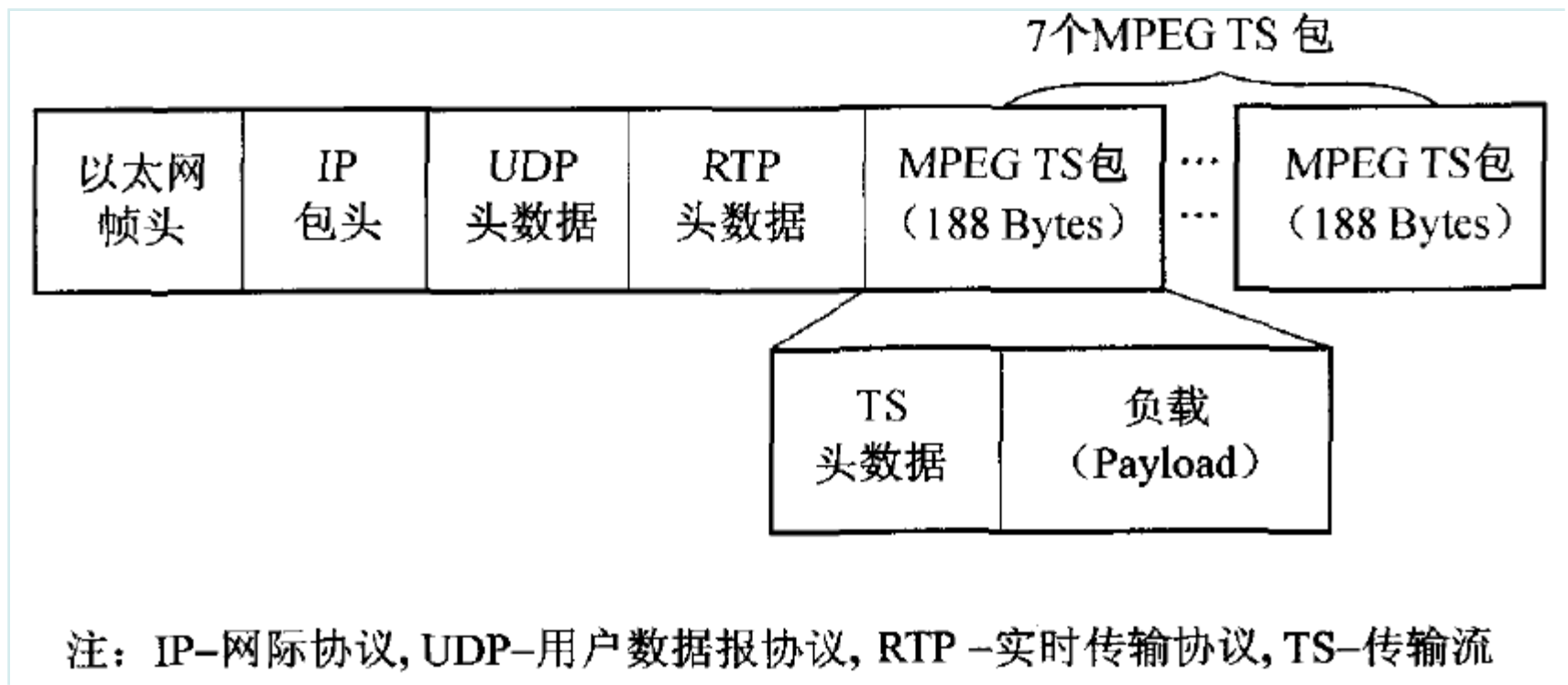




丢包与MPEG2视频质量关系的分析

IP网络上视频数据的封装

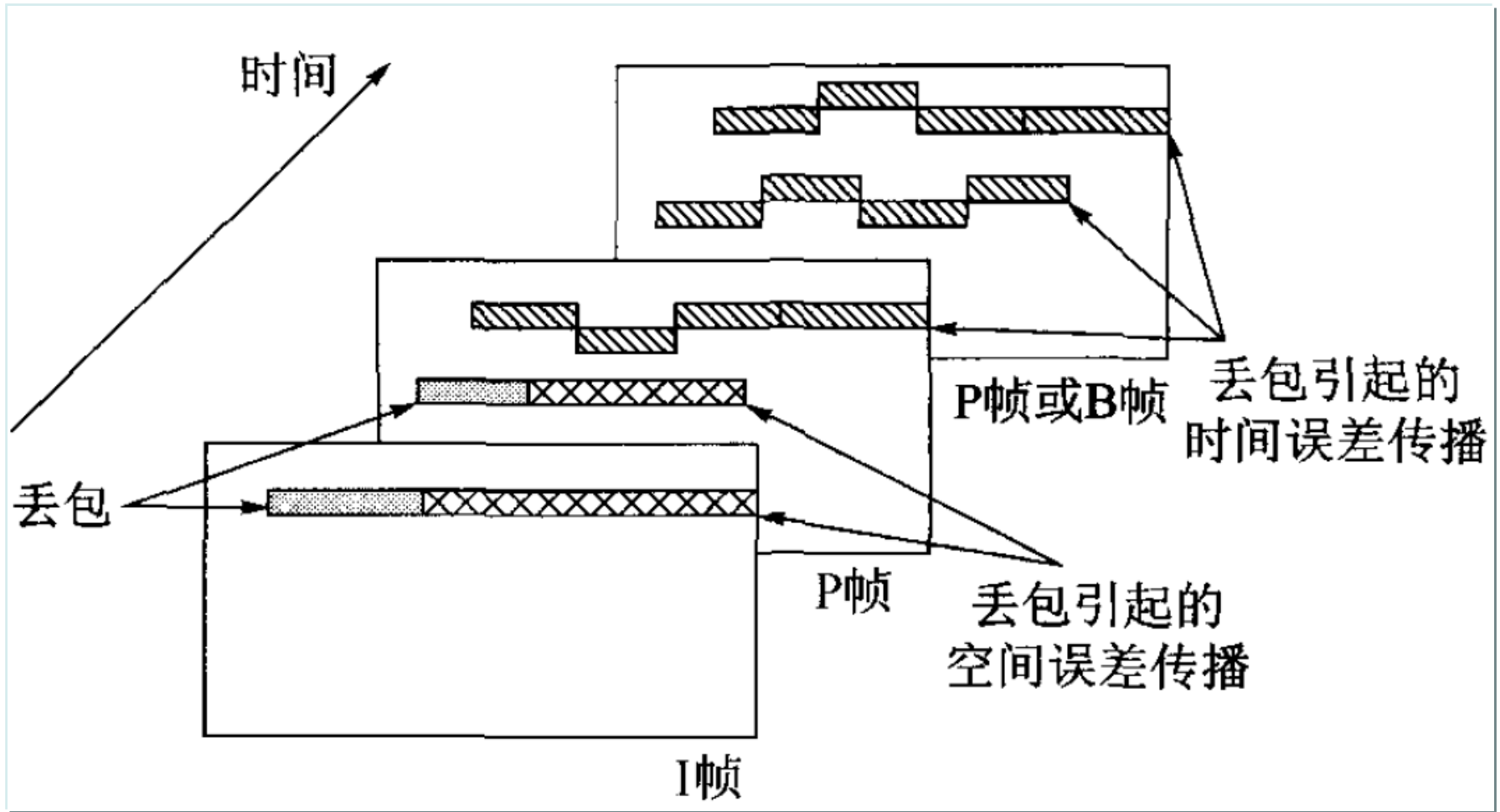
◆ UDP丢/错;MPEG TS块错





丢包与MPEG2视频质量关系的分析

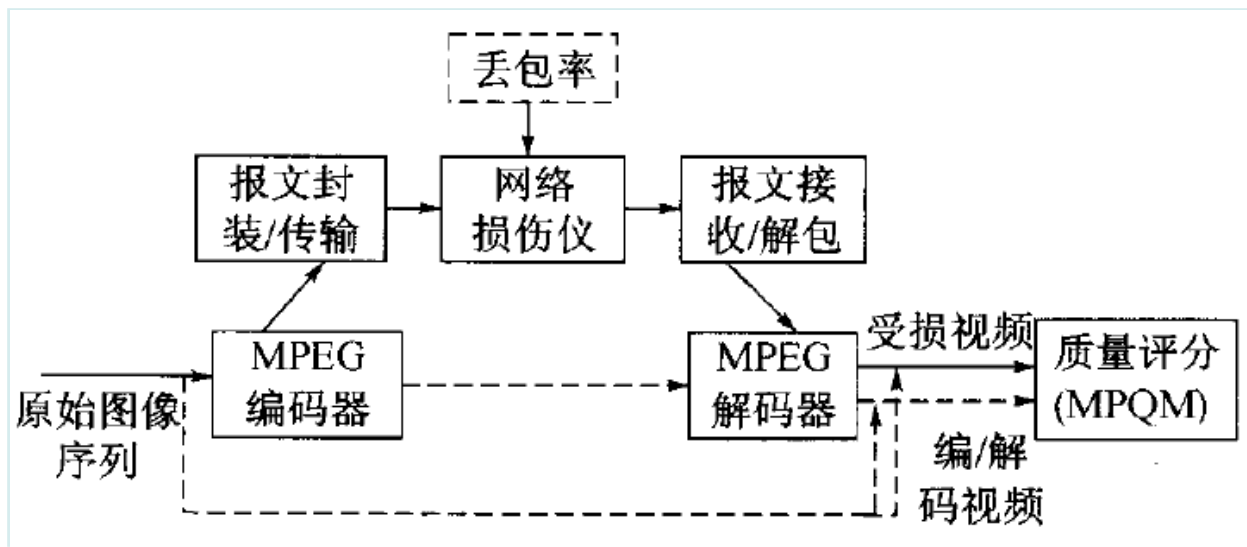
MPEG2丢包引起的错误传播





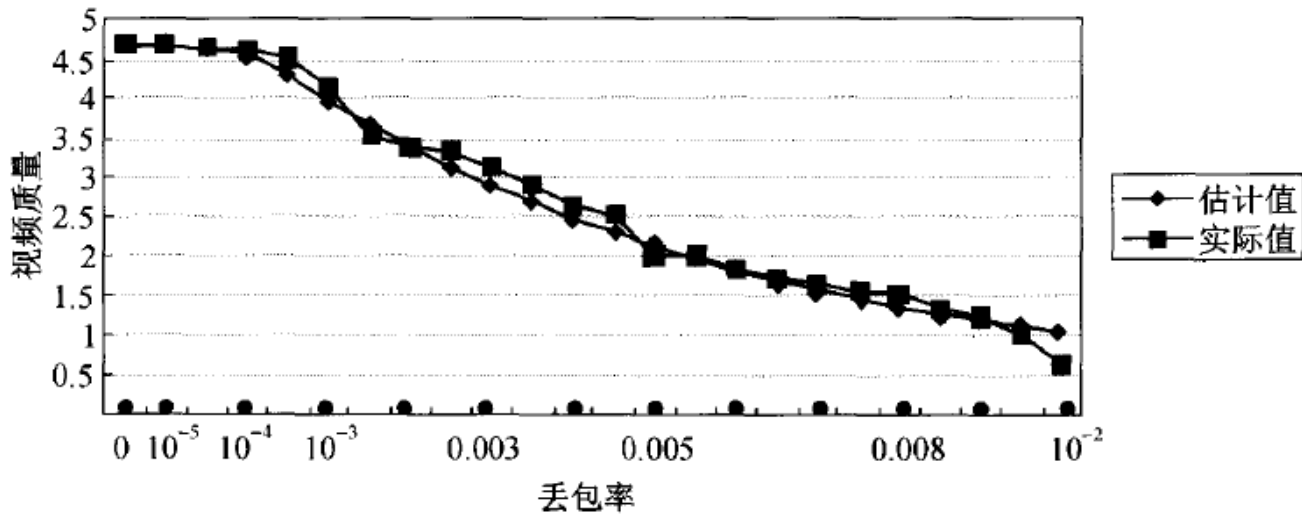
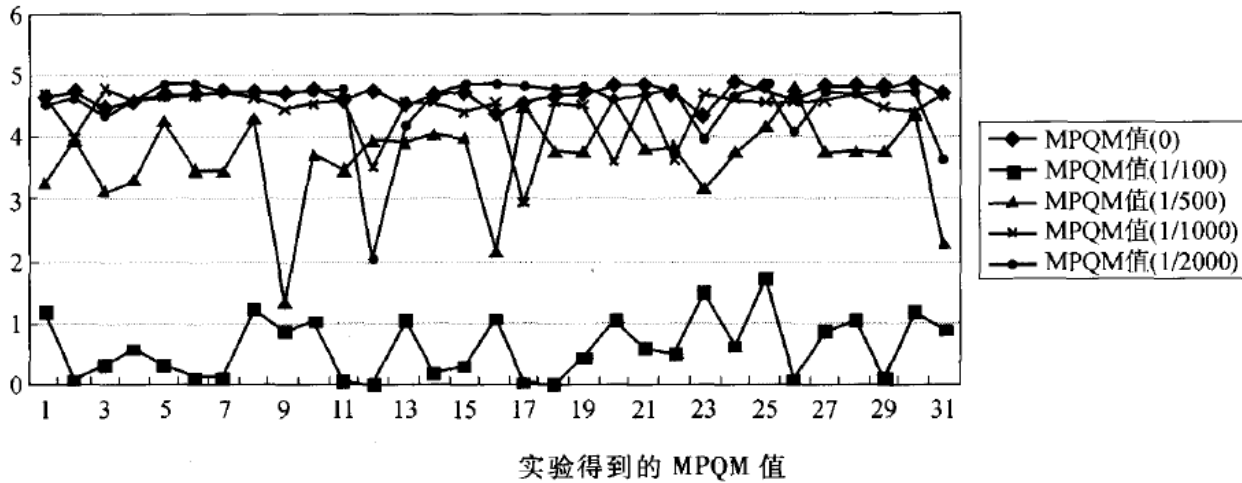
网络视频质量评估的基本模型

◆首先分析原始图像序列不经过网络传输而直接编解码后的视频质量的分值 Q_0 ，再分析它经网络传输后的视频质量情况；然后通过模拟不同的网络损伤来得出对应的视频质量分值 Q_1 、 Q_2 、...；最后把它们与根据网络视频质量评估模型得出的视频质量分值进行比较。



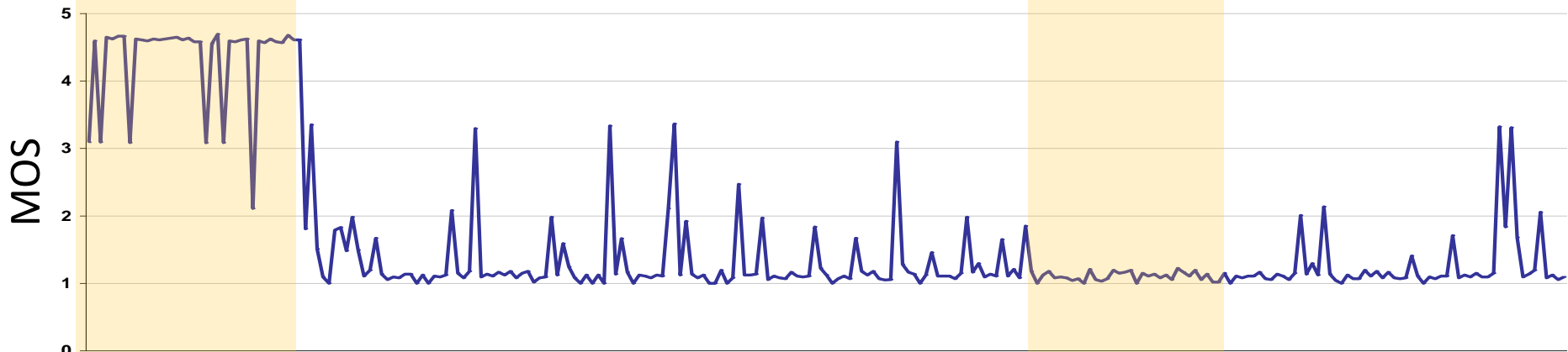


MOS分值与丢包率的关系





丢包率和QoE



Jitter 1 to 50ms	Packet Drop 1 in 500 periodic	Packet Drop 1 in 500 poisson	Packet Drop 1 in 500 uniform	Packet Drop 1 in 500 and Jitter	Duplicate Packets 1 in 10	Duplicate Packets 1 in 500	Reordering Packets 1 in 500
---------------------	----------------------------------	---------------------------------	---------------------------------	------------------------------------	------------------------------	-------------------------------	--------------------------------



Source <http://qoe.symmetricom.com> Sep 2008

{ynh, cxh}@ustc.edu.cn



视频传输系统质量分析方法

◆ Rate-Distortion Function

- 经典信息论

◆ Distortion -Distortion Function

- 1999

◆ Transmission Distortion模型

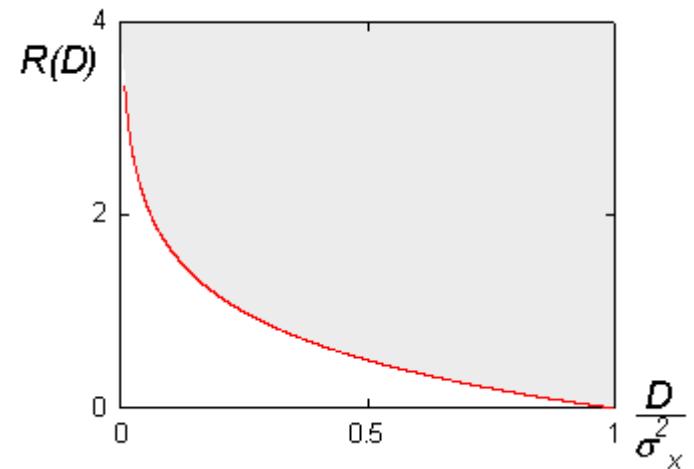
- 2006, Transmission Distortion Analysis for Real-Time Video Encoding and Streaming Over Wireless Networks
- 2011, Prediction of Transmission Distortion for Wireless Video Communication: Analysis



Rate-Distortion Function

◆ If we assume that $P_X(x)$ is Gaussian with variance σ^2 , and if we assume that successive samples of the signal X are stochastically independent (or, if you like, the source is *memoryless*, or the signal is *uncorrelated*), we find the following analytical expression for the rate–distortion function:

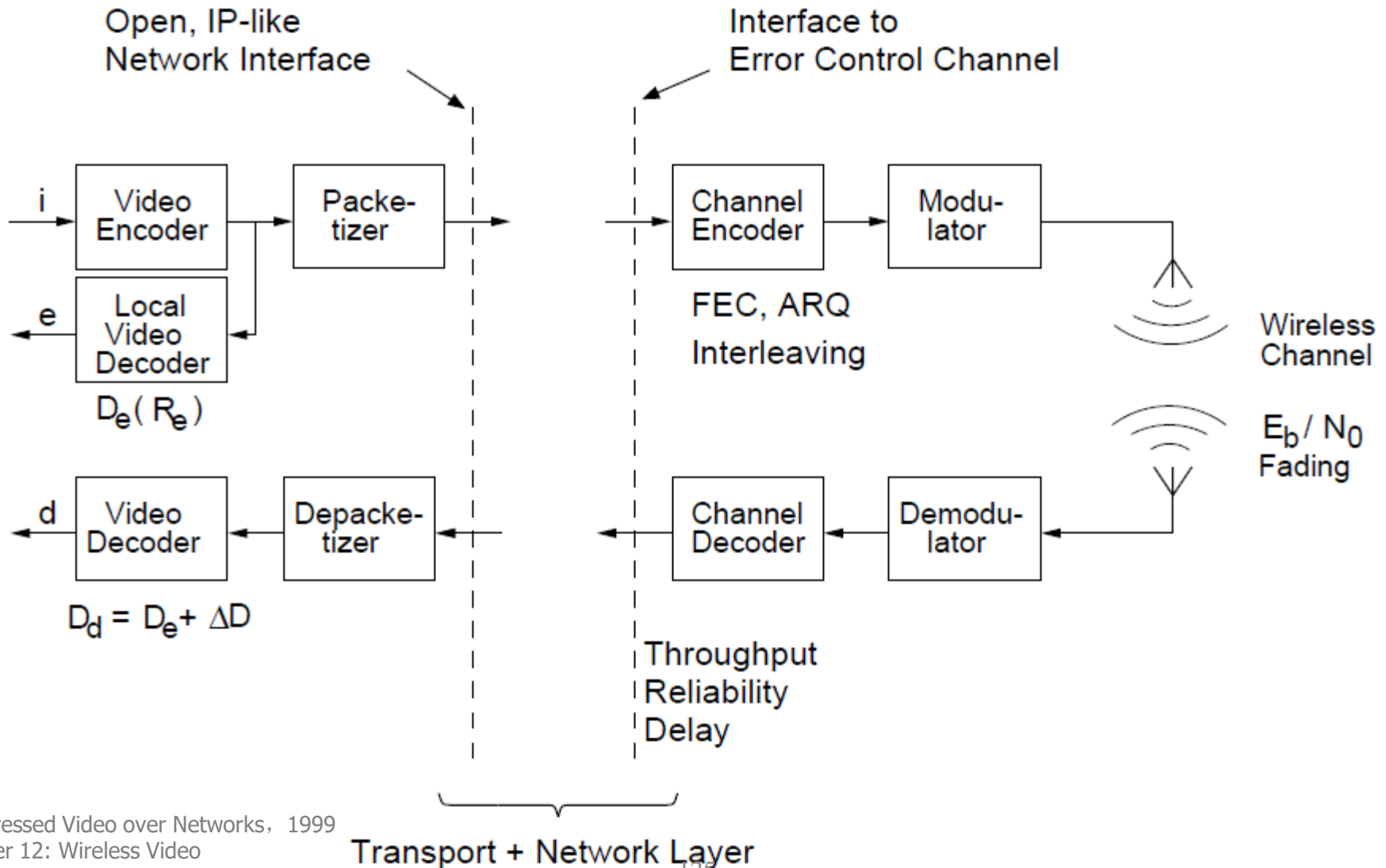
$$R(D) = \begin{cases} \frac{1}{2} \log_2(\sigma_x^2/D), & \text{if } D \leq \sigma_x^2 \\ 0, & \text{if } D > \sigma_x^2 \end{cases}$$





Distortion -Distortion Function

视频传输系统的基本构成





Distortion -Distortion Function

Distortion的度量

◆ 编码端失真

$$D_e = \frac{1}{XYT} \sum_{x=1}^X \sum_{y=1}^Y \sum_{t=1}^T (i[x, y, t] - e[x, y, t])^2$$

◆ 解码端失真

$$D_d = \frac{1}{XYTL} \sum_{x=1}^X \sum_{y=1}^Y \sum_{t=1}^T \sum_{l=1}^L (i[x, y, t] - d_l[x, y, t])^2$$

◆ 传输失真

$$\Delta D = D_d - D_e$$



Distortion -Distortion Function

用PSNR度量传输错误引起的Distortion

$$PSNR_e = 10 \log_{10} \frac{255^2}{D_e}$$

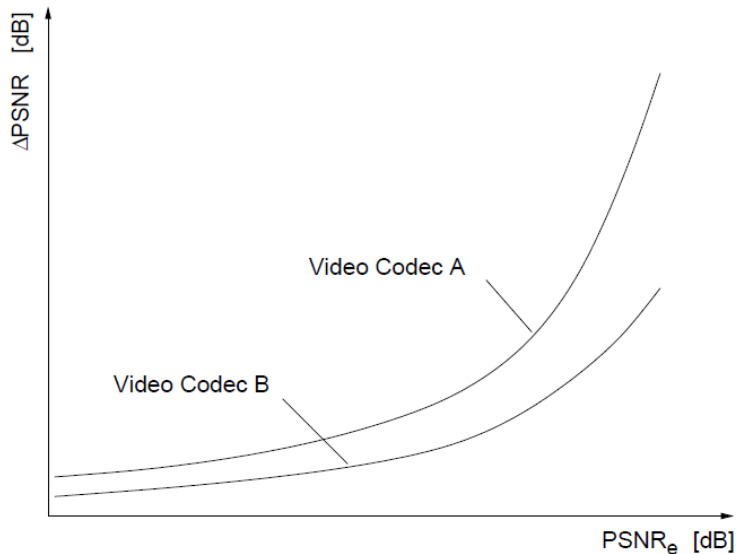
$$PSNR_d = 10 \log_{10} \frac{255^2}{D_d}$$

$$\Delta PSNR = PSNR_e - PSNR_d = 10 \log_{10} \frac{D_e}{D_d} = 10 \log_{10} \frac{D_e}{D_e + \Delta D}$$



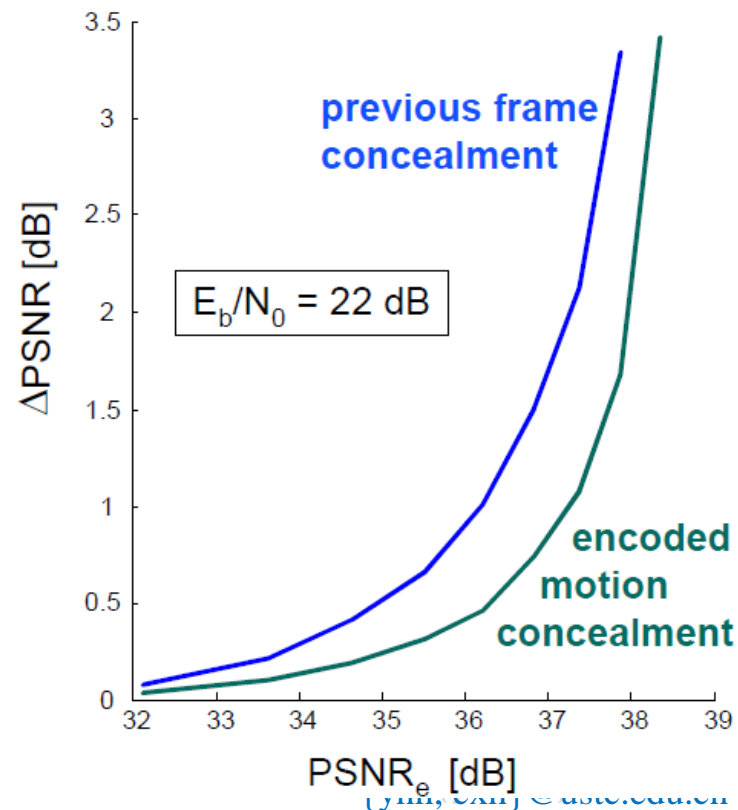
Distortion -Distortion Function

基于DDF的传输质量评估



不同编码器的DDF

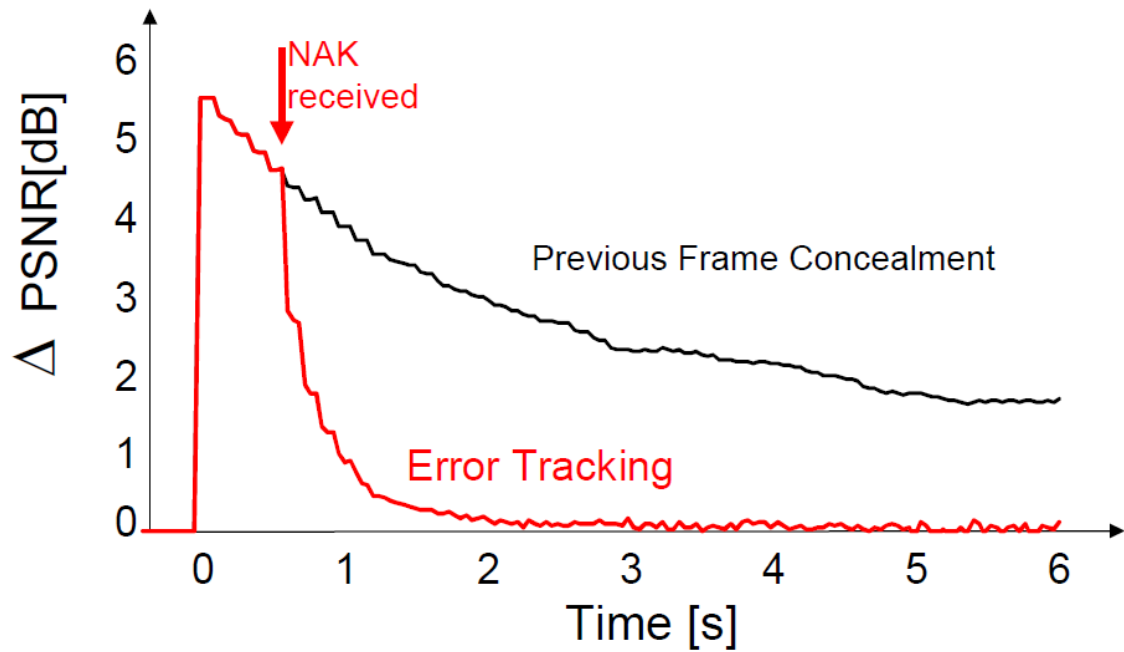
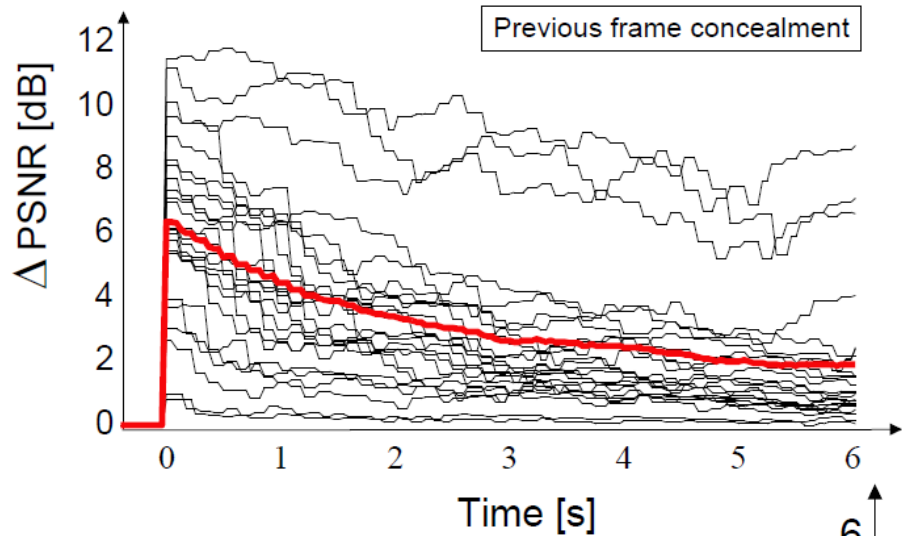
不同Error Concealment的DDF





Distortion -Distortion Function

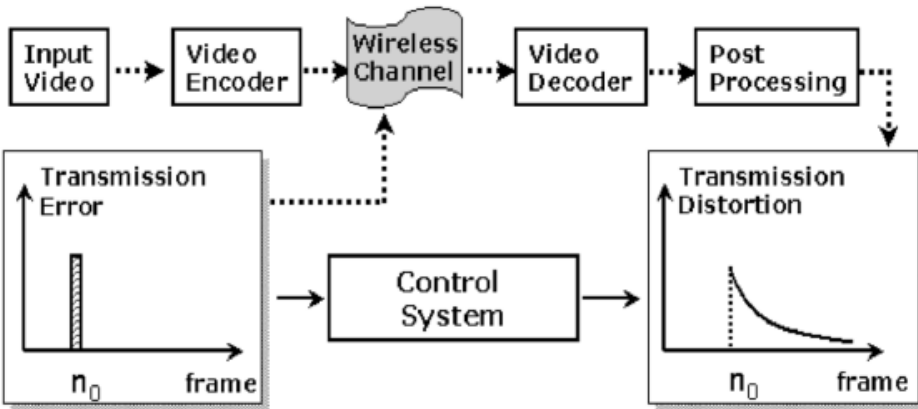
从突发错误中恢复



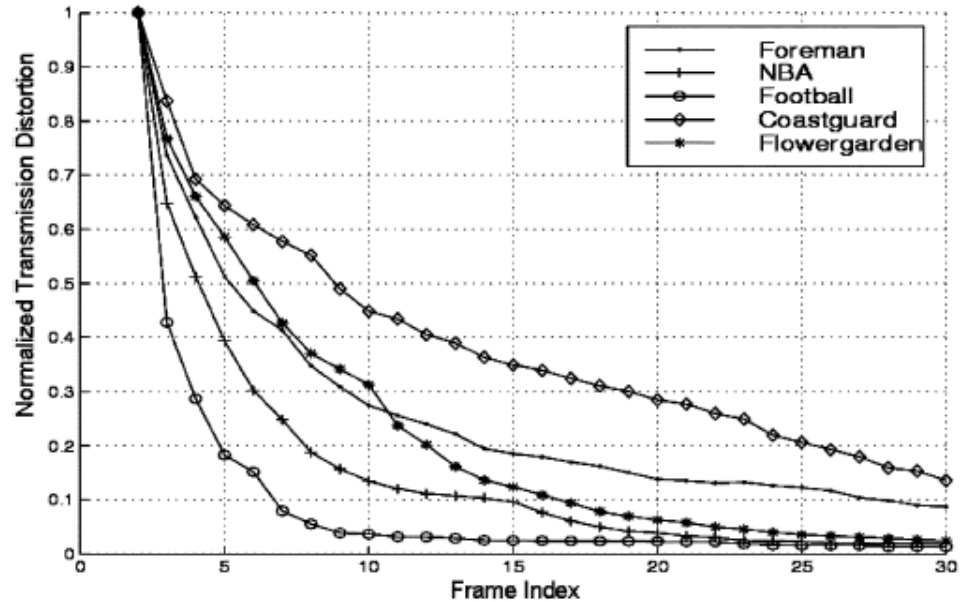


Transmission Distortion模型

传输失真的理论模型之一



$$\mathcal{H}(n) = D_t(n_0)e^{-\alpha(n-n_0)}$$



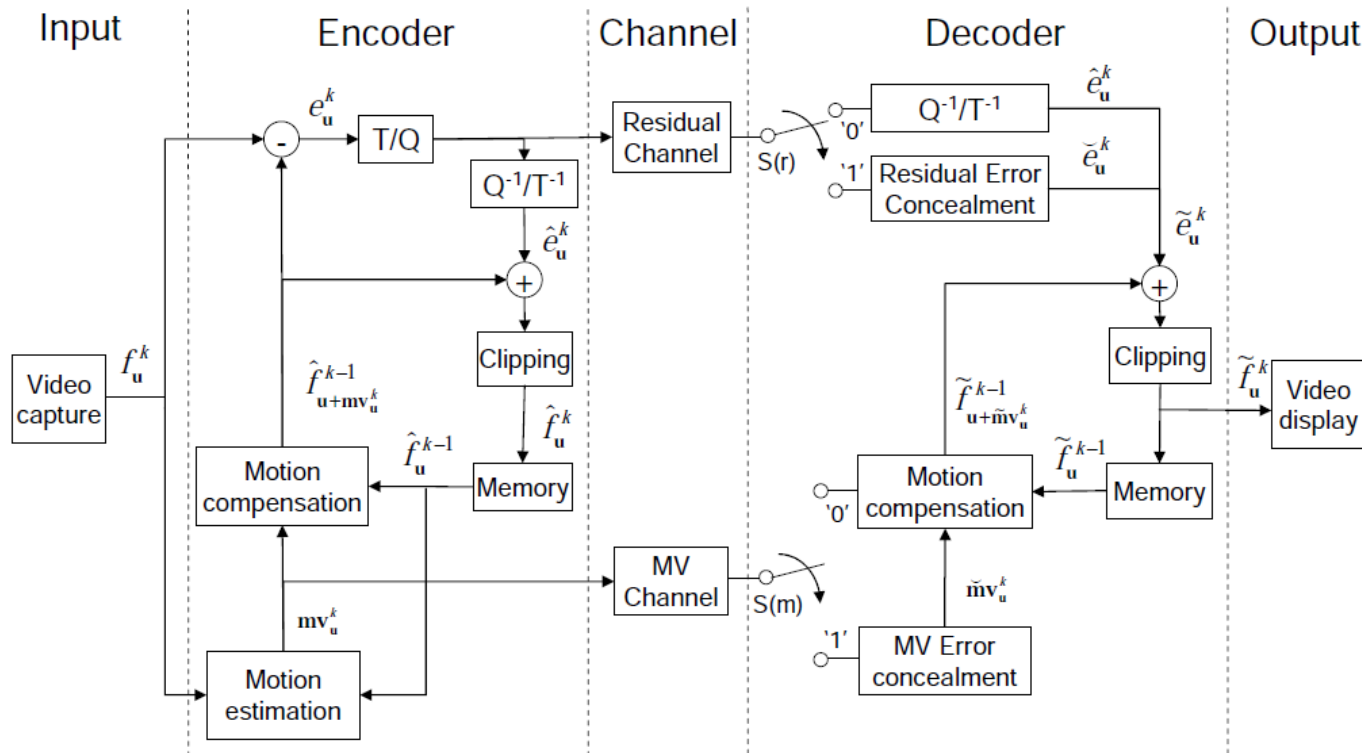


Transmission Distortion模型

传输失真的理论模型之二

derive the transmission distortion formula as a **closed-form function** of video frame statistics, channel error statistics, and system parameters.

$$D^k = D^k(r) + \lambda^k \cdot D^k(m) + \bar{P}^k(r) \cdot D^{k-1} + (1 - \bar{P}^k(r)) \cdot D^k(p) \cdot (1 - \beta^k)$$





近期视频传输失真的研究



<http://www.wu.ece.ufl.edu/>

Department of Electrical and Computer Engineering, University of Florida

Q. Chen, D. Wu, "**Delay-Rate-Distortion Model for Real-Time Video Communication**," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 8, pp. 1376--1394, August 2015.

J. Xu, D. Wu, "**Gamma Rate Theory for Causal Rate Control in Source Coding and Video Coding**," Journal of Visual Communication and Image Representation, vol. 25, no. 8, pp. 1886--1893, November 2014.

C. Li, D. Wu, H. Xiong, "**Delay-Power-Rate-Distortion Model for Wireless Video Communication Under Delay and Energy Constraints**," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 7, pp. 1170--1183, July 2014.

Z. Chen, P. Pahalawatta, A. M. Tourapis, D. Wu, "**Improved Estimation of Transmission Distortion for Error-resilient Video Coding**," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 4, pp. 636--647, April 2012.

MC 通用测试视频序列

◆ <http://trace.eas.asu.edu/yuv/index.html>



MC 小结：图像和视频质量评估

- ◆ 图像和视频的失真
- ◆ 质量评估的经典方法
 - 基于PSNR和MSE
 - 基于HVS (human visual system) 生理特征
 - 基于结构失真
- ◆ 网络视频
 - 简单测量：丢包率，包时延？
 - 理论方法：RDF、DDF, $H[n]$

