



UHD World Association
世界超高清视频产业联盟

世界超高清视频产业联盟标准

T/UWA 015—2022

高动态范围视频图像质量 第1部分：主观评价

Picture quality for high dynamic range video——Part 1: Subjective assessment

2022 - 03 - 30 发布

2022 - 03 - 20 实施

世界超高清视频产业联盟

目 次

前言.....	III
引言.....	V
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 主观评价要求.....	1
4.1 测试条件.....	1
4.2 信号源.....	2
4.3 测试素材选择.....	2
4.4 评价人员.....	3
4.5 评分方法.....	3
5 测试与评价.....	5
5.1 清晰度.....	5
5.2 图像噪声.....	5
5.3 白平衡.....	5
5.4 灰阶表现.....	5
5.5 色彩饱和度.....	5
5.6 色彩准确性.....	6
5.7 图像对比度.....	6
5.8 运动效果.....	6
5.9 高色域效果.....	6
5.10 峰值亮度效果.....	6
5.11 肤色效果.....	6
6 评价方法.....	6
6.1 评价流程.....	6
6.2 数据处理.....	7
6.3 结果的表达.....	8
6.4 主观评价报告的内容.....	8
附录 A（规范性）评价结果的筛选.....	9
附录 B（规范性）评价项目权重及分值表.....	10

前 言

本文件按照 GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是UWA标准《高动态范围视频图像质量》的第1部分。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由世界超高清视频产业联盟提出并归口。

本文件起草单位：中国电子技术标准化研究院、海信视像科技股份有限公司、京东方科技集团股份有限公司、小米通讯技术有限公司、北京数字电视国家工程实验室有限公司、北京淳中科技股份有限公司、深圳赛西信息技术有限公司、国家数字音视频及多媒体产品质量监督检验中心、虚拟现实/增强现实国家产品质量监督检验中心、康佳集团股份有限公司、厦门厦华科技有限公司、工业和信息化部电子第五研究所、上海索广映像有限公司、索尼（中国）有限公司、荣耀终端有限公司、广州视源电子科技有限公司

本文件主要起草人：孙婧、史培宁、王伟、郑晓航、顿胜堡、魏于凡、徐晖、贾凯、袁元春、罗少锋、孔令术、朱正元、韦胜钰、刘毅、赵晓莺、张冉、柳华玉、吴术霞、黄惠贞、陆凯平。

引 言

《高动态范围视频图像质量》由两部分构成。

- 第1部分：主观评价。
- 第2部分：客观评价。

全国团体标准信息平台

高动态范围视频图像质量 第1部分：主观评价

1 范围

本文件规定了高动态范围显示终端视频图像质量的主观评价要求和对应的主观评价方法。
本文件适用于高动态范围电视机和高动态范围显示器。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GY/T 315 高动态范围电视节目制作和交换图像参数值

SJ/T 11324 数字电视接收设备术语

SJ/T 11746-2019 超高清晰度电视机显示性能测试方法

3 术语和定义

SJ/T 11324 界定的以及下列术语和定义适用于本文件。

3.1

主观评价 subjective assessment

评价人员对被测显示终端视频图像质量的主观反应。

3.2

测试素材 test fodder

用于图像质量的主观评价，其内容特定要求有静止图像、活动图像或这两种类型组合的图像序列。

3.3

评级标度 grading scale

主观评价中用于评定图像质量的评定尺度。

3.4

伪轮廓现象 false contour

平滑过渡的图像出现了可见的灰度轮廓。

3.5

高动态范围 high dynamic range

HDR

在现有动态范围基础上，拓展高亮和低亮动态范围，增加画面高亮和低亮图像细节。

4 主观评价要求

4.1 测试条件

主观评价环境以及观看条件应符合表1的规定。

表 1 观看条件

项目	参考值
背景 ^a	D 65中性灰
暗环境背景亮度	$\leq 5 \text{ cd/m}^2$
亮环境照度	200 lx
室内环境光照明	避免光线落在屏幕上
观看距离	对于3840×2160格式：1.6~3.2倍图像高度 对于7680×4320格式：0.8~1.6倍图像高度
评价人员的座位安排	4 K：水平方向在中垂线±30° 以内 8 K：水平方向在中垂线±45° 以内
电源电压	220(1±10%) V, 频率振动 $\leq \pm 2 \%$, 谐波分量 $\leq \pm 5 \%$

^a “背景” 仅针对暗环境。

4.2 信号源

信号源产生的测试信号应符合GY/T 315规定，输出接口宜采用数字视频接口，如采用其他接口测试，应记录所采用的接口类型。信号的格式应符合表2的要求。

表 2 图像格式

图像格式	扫描方式	场频 Hz	幅型比	编码比特
1920×1080	逐行	50、100、120	16: 9	10
3840×2160				
7680×4320				

4.3 测试素材选择

选择主观评价测试素材应符合以下要求：

- 测试图像的色彩、形式、细节、亮度和对比度等不应使评价人员分心；
- 测试图像通常包含一定数目的静止、快速及慢速运动物体的图像序列，应包含不同的运动形式，如平移、推拉或旋转，及不同运动方向，如横向、竖向或斜向；
- 所选的测试图像内容应对不同损伤具有一定的敏感程度，如静止图像纹理、动态图像纹理、亮度再现、色彩再现和运动再现等具有敏感特性的评价因素；
- 测试图像应包含不同皮肤的人物、蓝天、绿地、红旗等图案的一种或其组合；

- e) 测试图像应包括不同的平均亮度、对比度，不同的细节丰富程度以及不同的画面构图。例如包含室内和室外场景，包含远景、中景和近景、包含人物和景色、包含多细节、中等细节和少细节、包含背景亮和暗等因素；
- f) 测试图像应包含多种色调和色彩饱和层次；
- g) 测试图像应包含支持不同色域空间的内容,如：支持 BT.709 色域空间的内容、支持 BT.2020 色域空间的内容等。

4.4 评价人员

评价人员应具有正常的视力（含矫正视力）和色彩视觉，应熟练的掌握评价方法，具有独立的分析判断能力。

评价人员对于测试系统引入的图像不具有专长，所需评价人员的数目取决于测试程序的感受度和信度，应至少20位评价人员组成。

4.5 评分方法

4.5.1 刺激比较法

4.5.1.1 测试系统组成

测试系统的组成见图1。对比样机应与被测样机具备相同的技术参数，如：相同显示形态、相同的物理分辨率和相同的屏幕尺寸等。

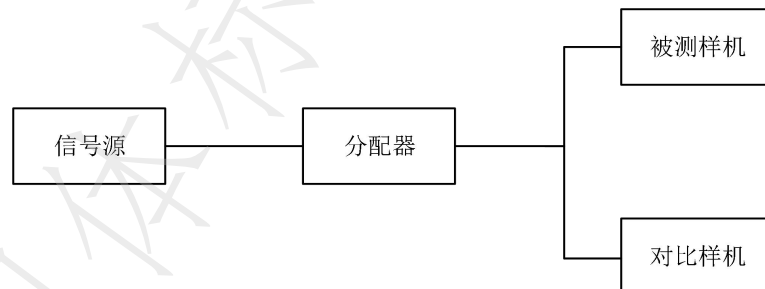


图 1 测试系统框图——刺激比较法

4.5.1.2 评分标度

刺激比较法使用连续量表法判断，每个评价人员在垂直标尺上做标记，对每次演示的总体图像质量进行评价，见表3。为了避免量化误差，标度提供了连续的等级系统，并将垂直标尺分成了相等的七段长度，与比较量表相对应，见图2。

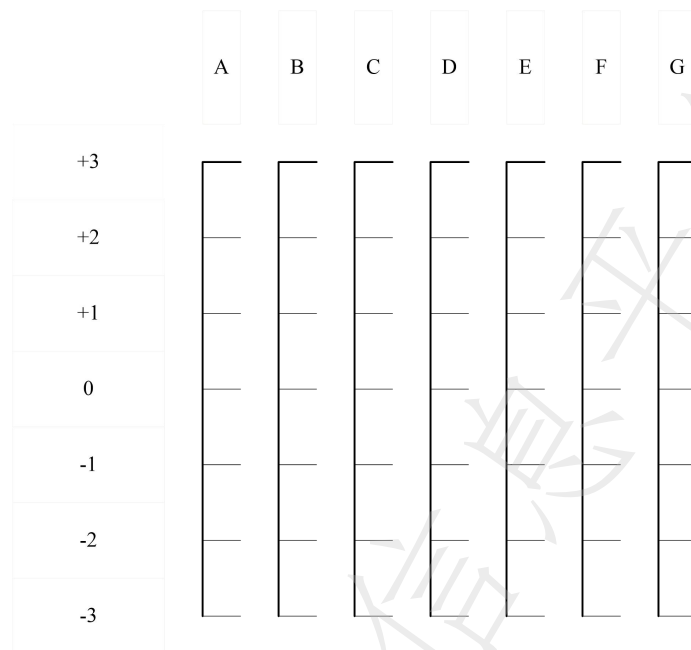


图2 刺激比较连续质量标度

表 3 评级标度描述

等级	描述
-3	差得多
-2	差
-1	稍差
0	相同
1	稍好
2	好
3	好得多

4.5.2 单刺激法

4.5.2.1 测试系统组成

测试系统的组成见图3。

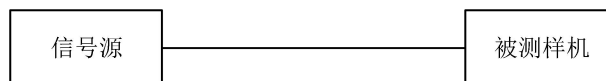


图 3 测试系统框图——单刺激法

4.5.2.2 评分标度

单刺激法使用数值量表法判断，评价人员为每一个图像或图像序列打分，分值范围为0~100。分值可与五级质量标度相对应，见表4。

表 4 分值与五级质量量表对应关系

分值	等级	质量术语描述
100	5	优
80	4	良
60	3	中
40	2	差
20	1	劣

5 测试与评价

5.1 清晰度

应选择可反映图像细节程度的图像序列进行评价,评价序列应包含静态图像序列和动态图像序列两种。静态被测序列观察如发丝、动物茸毛、树枝、灌木丛、猪肉纹理等场景,动态被测序列观察如行驶过程中的汽车、球场上跑动的人的面部特征、移动场景中的建筑等,按照各细部纹理和边界的清晰程度给予评价。

5.2 图像噪声

应选择可反映噪声效果的图像序列进行评价,评价的序列应含静态图像序列和动态图像序列两种。观察被测序列如:人脸、头发、衣服、建筑墙、背景等静止和运动的场景,按照各部噪声令人厌恶的程度给予综合评价。

5.3 白平衡

应选择可反映不同灰度等级的图像序列进行评价,评价的序列应含静态图像序列和动态图像序列两种。静态序列可使用灰场、全白场等特定图卡观察,动态被测序列观察如白色餐具、夜景等场景,按照各部灰阶还原程度给予综合评价。

5.4 灰阶表现

应选择可反映不同灰度丰富程度的图像序列进行评价,评价的序列应含静态图像序列和动态图像序列两种。静态序列可选取特定图卡观察灰阶等级呈现能力,如不同等级的灰阶图卡;动态被测序列观察如夜空、窗帘、门板等场景,按照各部灰阶等级呈现能力以及伪轮廓轻重程度给予综合评价。

5.5 色彩饱和度

应选择反映不同色彩饱和层次的图像序列进行评价,评价的序列应含静态图像序列和动态图像序列两种。静态序列可使用多种颜色种类、不同饱和度的色阶序列进行观察,动态被测序列观察如红花、果蔬、油菜花等场景,按照各部层次表现力以及色彩鲜艳程度给予综合评价。

5.6 色彩准确性

应选择具有丰富色彩色调的图像序列进行评价，评价的序列应含静态图像序列和动态图像序列两种。观察被测序列中如：不同肤色的人群、蓝天、绿地、红旗等场景，按照各部色彩还原能力给予综合评价。

注：色彩准确性在不同的文件中也叫色差。

5.7 图像对比度

应选择反映不同对比度以及细节丰富程度的图像序列进行评价。观察被测序列中如：动态灰阶、太阳升起的过程、阳光照射下的体育场等场景，按照各部亮暗场层次展现程度、图像整体的纵深感给予综合评价。

5.8 运动效果

应选择反映不同运动速度和运动形式的图像序列进行评价。观察被测序列中如：横向和纵向切换场景、快速运动的足球、飞过高楼的滑翔机等场景，按照整体运动流畅程度、运动物体边缘轮廓以及是否有破碎、边界保护效果等各方面给予综合评价。

5.9 高色域效果

应选择支持不同色域空间的图像序列进行评价，评价的序列应含静态图像序列和动态图像序列两种。静态序列可选取特定图卡观察色彩再现能力，如色彩过渡图；动态被测序列观察如红玫瑰、绿色丛林、阳光照射下的广告图等场景，按照各部色彩表现能力给予综合评价。

5.10 峰值亮度效果

应选择高对比度图像序列进行评价。此被测序列中应包含亮、暗交替出现的场景，观察被测序列中如：璀璨的星空、夜晚灯光秀、暗背景电影片头亮字幕、从黑山洞走到阳光下海滩等场景，按照高亮部分的明暗程度以及同一序列不同场景切换前后亮暗程度给予评价。

5.11 肤色效果

应选择不同肤色、不同场景包含人物的图像序列进行评价，评价的序列应含静态图像序列和动态图像序列两种。静态序列可选取不同国籍的人脸，动态序列可选战争片夜景中的人物、新闻联播主播、NBA赛事外国运动员等场景，按照肤色效果给予评价。

6 评价方法

6.1 评价流程

6.1.1 标准工作状态调整

除另有规定外，被测显示终端的状态按SJ/T 11746-2019中4.4.1中a)、b)和d)的规定进行测试状态调整。

6.1.2 稳定时间

被测显示终端的稳定时间按SJ/T 11746-2019中4.4.4中规定执行。

6.1.3 评价过程

6.1.3.1 评价训练

在每一评价训练开始时，应向评价人员详细介绍评价方法、容易产生的质量损伤类型、图像显示顺序和时间、评分制、评分等级好使用设备等。介绍人员语气中不应带有影响评定的暗示。

评价训练阶段应持续半小时以内。第一段开始时，应播放5个左右“模拟演示”，以稳定评价人员的评分。这几个序列中给出的数据不能在测试结果中考虑。如果需要多个测试阶段，在后续阶段开始时仅需要3个左右的“模拟演示”。

训练序列播放顺序采用随机顺序，训练序列不应使用正式的测试序列，但应与正式的测试序列具有可比性。

6.1.3.2 正式评价

评价人员的评价过程要求如下：

- a) 测试序列应按照评价表中的维度顺序播放，测试阶段每一个测试序列演示3次。第一次演示用于稳定观察者的意见，此次演示不评分；对于测试序列的评分是从第二个和第三个演示中得到的。序列演示以及评分顺序如图4所示。

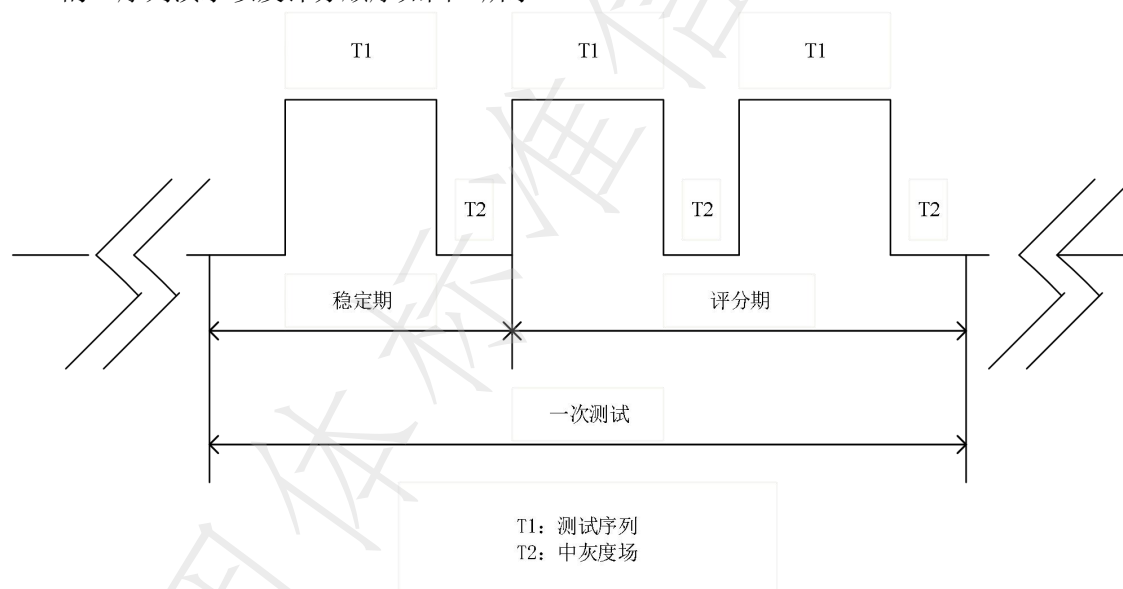


图4 序列演示和评分顺序

- b) 评价人员应按有关评分等级的说明对被测图像独立地进行评价打分，试验过程中评价人员应集中注意力，防止环境干扰对评价人员的视觉产生影响。

6.2 数据处理

刺激比较法需要先将结果归一化为0和100之间的整数。

评价人员评价结果的筛选见附录A。

刺激比较法以及单刺激法都需要计算出每个被测序列的平均分 \bar{u}_{jkr} ，见公式(1)：

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr} \dots \dots \dots (1)$$

式中：

u_{ijkr} ——评价人员*i*在*j*、序列/图像*k*、重复*r*次情况下的评分；

N ——评价人员的数目；

j ——测试条件的数目，包括基准在内；

k ——测试图像或序列的数目；

r ——重复次数。

每一个被测序列的标准偏差 S_{jkr} 由式（2）给出：

$$S_{jkr} = \sqrt{\frac{\sum_{i=1}^N (\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}} \dots\dots\dots (2)$$

δ_{jkr} 代表置信区间的半径微变量，由式（3）给出：

$$\delta_{jkr} = 1.96 \times \frac{S_{jkr}}{\sqrt{N}} \dots\dots\dots (3)$$

式中：

S_{jkr} ——每一个被测序列的标准偏差；

N ——评价人员的数目。

95 %置信度区间为：

$$[\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} + \delta_{jkr}]$$

6.3 结果的表达

针对每个被测序列的平均分 \bar{u}_{jkr} ，计算测试条件 j 时的平均分 \bar{u}_j ，最后应按照附录B进行评分汇总，得出显示终端终评得分 U 。

6.4 主观评价报告的内容

报告中还应包含以下内容：

- a) 被测显示终端型号、版本等信息；
- b) 评价使用的设备以及测试素材的情况；
- c) 主观评价采用的方法；
- d) 评价人员人数；
- e) 显示终端终评得分，其中单刺激法终评得分为数据处理后的结果；刺激比较法中对比样机按照50分终评分参与计算，若对比样机图像质量实得分为 U_{ds} ，则被测样机加权后的终评得分 U_z 可按照式（4）得出：

$$U_z = U \times \frac{U_{ds}}{50} \dots\dots\dots (4)$$

- f) 总视频质量的描述。

附 录 A
(规范性)
评价结果的筛选

首先,利用 β_2 (通过计算函数的峰态系数,即四阶中心距与二阶中心距平方的比值)确定每个测试显示的得分是否是正态分布。对于每个测试显示,计算平均分 \bar{u}_{jkr} 、标准偏差 S_{jkr} 和峰度系数 β_{2jkr} , β_{2jkr} 由式(A.1)给出。

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \quad \text{其中 } m_x = \frac{\sum_{i=1}^N (u_{ijkr} - \bar{u}_{jkr})^x}{N} \dots\dots\dots (A.1)$$

式中:

m ——中间过程量;

x ——取值为2或4;

N ——评价人员的数目;

j ——测试条件的数目,包括基准在内;

k ——测试图像或序列的数目;

r ——重复次数。

对于每一位评价人员 i ,找出每一 P_i 和 Q_i ,即:

若 $2 \leq \beta_{2jkr} \leq 4$,这个分布可以取做正态(即视为正常)。则:

$$\text{当 } u_{ijkr} \geq \bar{u}_{jkr} + 2 S_{jkr} \text{ 时 } P_i = P_i + 1;$$

$$\text{当 } u_{ijkr} \leq \bar{u}_{jkr} - 2 S_{jkr} \text{ 时 } Q_i = Q_i + 1$$

若 $\beta_{2jkr} < 2$ 或者 $\beta_{2jkr} > 4$,则:

$$\text{当 } u_{ijkr} \geq \bar{u}_{jkr} + \sqrt{20} S_{jkr} \text{ 时 } P_i = P_i + 1;$$

$$\text{当 } u_{ijkr} \leq \bar{u}_{jkr} - \sqrt{20} S_{jkr} \text{ 时 } Q_i = Q_i + 1$$

最后,如果满足下面的计算结果,则剔除评价人员 i 。

$$\frac{(P_i + Q_i)}{j \cdot k \cdot r} > 0.05, \text{ 且 } \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3$$

式中:

u_{ijkr} ——评价人员 i 的评分。

附 录 B
(规范性)
评价项目权重及分值表

表B.1规定了各评价项目对应的权重及分值。

表 B.1 评价项目权重及分值表

序号	评价项目 ^a	权重系数	加权满分值
1	清晰度	15 %	15
2	图像噪声	10 %	10
3	白平衡	3 %	3
4	灰阶表现力	8 %	8
5	色彩饱和度	8 %	8
6	色彩准确性	8 %	8
7	图像对比度	15 %	15
8	运动效果	10 %	10
9	高色域效果	8 %	8
10	峰值亮度效果	8 %	8
11	肤色效果	7 %	7
	总评分	100 %	100
^a 此处“评价项目”即为文中“测试条件”。			