

# ITU-R BT.2163-0 建议书

(11/2023)

BT系列：广播业务（电视）

## 评价高动态范围电视亮度的客观测量算法



## 前言

无线电通信部门的作用是确保所有无线电通信业务，包括卫星业务，合理、公平、有效和经济地使用无线电频谱，并开展没有频率范围限制的研究，在此基础上通过建议书。

无线电通信部门制定规章制度和政策的职能由世界和区域无线电通信大会以及无线电通信全会完成，并得到各研究组的支持。

## 知识产权政策（IPR）

国际电联无线电通信部门（ITU-R）的 IPR 政策述于 ITU-R 第 1 号决议所参引的《ITU-T/ITU-R/ISO/IEC 的通用专利政策》。专利持有人用于提交专利声明和许可声明的表格可从 <http://www.itu.int/ITU-R/go/patents/zh> 获得，在此处也可获取《ITU-T/ITU-R/ISO/IEC 的通用专利政策实施指南》和 ITU-R 专利信息数据库。

### ITU-R 建议书系列

（可同时在以下网址获得：<http://www.itu.int/publ/R-REC/zh>）

系列	标题
BO	卫星传输
BR	用于制作、存档和播放的记录；用于电视的胶片
BS	广播业务（声音）
<b>BT</b>	<b>广播业务（电视）</b>
F	固定业务
M	移动、无线电测定、业余无线电以及相关卫星业务
P	无线电波传播
RA	射电天文
RS	遥感系统
S	卫星固定业务
SA	空间应用和气象
SF	卫星固定和固定业务系统之间频率共用和协调
SM	频谱管理
SNG	卫星新闻采集
TF	时间信号和标准频率发射
V	词汇和相关课题

注：本ITU-R建议书英文版已按ITU-R第1号决议规定的程序批准。

电子出版物  
2024年，日内瓦

© 国际电联 2024

版权所有。未经国际电联书面许可，不得以任何手段翻印本出版物的任何部分。

## ITU-R BT.2163-0建议书

评价高动态范围电视亮度的客观测量算法<sup>1,2</sup>

(ITU-R第142-3/6号课题)

(2023)

## 范围

本建议书规定了一种基于平均图像流明度确定图像等级的测量算法，该算法可用于评估单个图像的亮度。基于图像等级的其他指标是时间图像等级和图像等级响应，这些方法或可用于模拟人类视觉系统对图像序列的响应。

## 关键词

视频测量、亮度、电视制作、高动态范围、HDR、电视、HDR-TV、国际节目交换

国际电联无线电通信全会，

考虑到

- a) 高动态范围电视（HDR-TV）提供了非常宽的图像流明度等级动态范围；
- b) 亮度的大幅跃升可能会让观看者感到惊讶；
- c) 监测图像亮度有助于了解观看者的体验；
- d) 人类视觉系统的适应状态受图像的时间序列影响；
- e) ITU-R BT.2100建议书规定了HDR电视的两种形式，即感知量化（PQ）和混合对数伽马（HLG），预计将以这两种格式制作；
- f) HLG参考显示器通常使用1 000 cd/m<sup>2</sup>的标称峰值流明度；
- g) 参考周边环境条件在ITU-R BT.2100建议书表3中进行规定；
- h) 制作中采用的现有视频测量方法不提供表示主观图像亮度的数值；
- i) 为了便于评价节目的主观视频亮度，标准化的测量方法是有用的，

认识到

使用这种算法将补充而不是取代传统的波形监测，

建议

- 1 当需要测量根据ITU-R BT.2100建议书制作的高动态范围图像的主观亮度以促进节目制作和交换时，可使用第1节中规定的图像等级算法；

---

<sup>1</sup> 本建议书并不能取代ITU-R BT.1702建议书中所载的指南，该建议书旨在保护观众中患有光敏性癫痫症的易感人群，这些人很容易因闪光（包括某些类型的闪光电视图像）而引发癫痫发作。

<sup>2</sup> 在对各种各样的运动图像进行测试后，以及在结合当前正在开发的计量规范进行测试时，可能需要对算法进行修订。

2 在第2段中规定的时间图像等级测量算法可用于评估图像序列对观看者适应水平的作用情况；

3 在第3段中规定的对眼睛响应情况、图像等级响应情况的近似可用于评价节目中亮度变化的重要性，

进一步建议

1 本建议书不应用作亮度调节工具，对运营商施加限制或者限制内容创作；

2 由于对运动图像算法的验证有限，因此ITU-R可能希望考虑对本建议书进行早期更新和改进。

## 1 图像等级测量算法

本节规定了图像等级（IL）测量算法，该算法可能有助于评估整体图像亮度。IL算法是第2和3段中本建议书中规定的其他指标的基础。

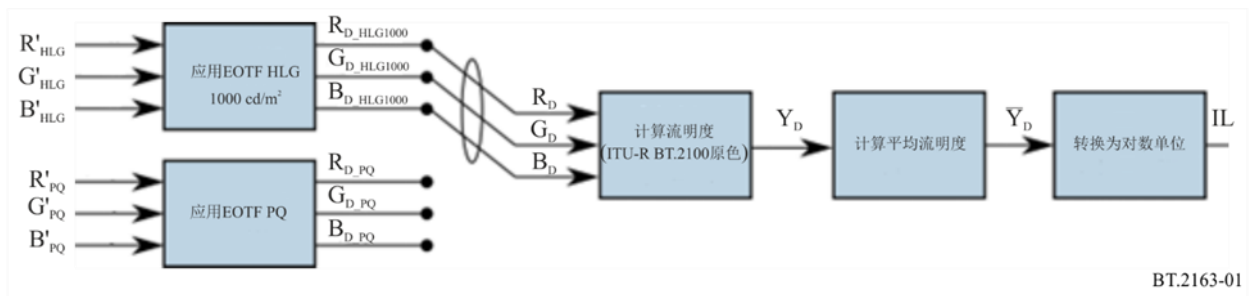
IL测量算法的输入是非线性的ITU-R BT.2100建议书HLG或PQ信号 $R'G'B'$ ，范围为[0:1]。

该算法包括四个阶段：

- 1 转换为线性显示光；
- 2 计算流明度分量；
- 3 计算平均显示流明度；
- 4 转换为对数单位。

图1显示了算法的框图。

图1  
显示图像等级测量算法的框图



### 1.1 转换为线性显示光

对于HLG信号 $R'G'B'_{HLG}$ ，应用ITU-R BT.2100建议书表5中规定的HLG EOTF，对于标称峰值流明度为1 000 cd/m<sup>2</sup>的显示器，使用ITU-R BT.2100建议书表5注释5f中规定的 $\gamma$ 值1.2。产生的信号是在[0:1 000] cd/m<sup>2</sup>范围内的显示线性光 $R_D G_D B_D_{HLG}$ 。

对于PQ信号 $R'G'B'_{PQ}$ ，应用ITU-R BT.2100建议书表4中规定的PQ EOTF。产生的信号是在 $[0:10\ 000]$  cd/m<sup>2</sup>范围内的显示线性光 $R_D G_D B_D$ 。

此后，分量线性显示光信号被称为 $R_D G_D B_D$ ，而不管该信号是源自HLG还是PQ。

## 1.2 流明度分量的计算

使用以下公式来计算线性显示流明度信号 $Y_D$ ：

$$Y_D = 0.2627R_D + 0.6780G_D + 0.0593B_D$$

在附件1中描述了一种近似方法。

## 1.3 平均显示流明度的计算

而后计算视频帧的平均显示流明度 $\bar{Y}_D$ ，单位为cd/m<sup>2</sup>。

$$\bar{Y}_D = \frac{1}{H \times V} \sum_{h=0}^{H-1} \sum_{v=0}^{V-1} Y_D(h, v)$$

其中：

$Y_D(h, v)$ ：水平位置 $h$ 、垂直位置 $v$ 上像素值的线性显示流明度；

$H$ ：水平像素数；

$V$ ：垂直像素数。

## 1.4 转换为对数单位

以cd/m<sup>2</sup>为单位的平均显示流明度转换为以2为底的对数单位，归一化值为1 cd/m<sup>2</sup>。

注 - 1 cd/m<sup>2</sup>的平均显示流明度产生的IL值为0。

$$IL = \log_2 \frac{\bar{Y}_D}{1}$$

在附件2中提供了有关该算法的背景研究资料。

## 2 时间图像等级算法

观看一系列图像会影响人类观察者的适应状态，并因此影响对图像等级的感知效果。时间图像等级（TIL）<sup>3</sup>可用于评估图像亮度随时间的变化。TIL计算如下：

$$TIL(t) = \begin{cases} IL(t) & \text{若 } t = 0 \\ TIL(t-1) \left(1 - \frac{1}{\tau+1}\right) + IL(t) \left(\frac{1}{\tau+1}\right) & \text{若 } t > 0 \end{cases}$$

其中， $t$ 是从0开始的帧数， $\tau$ 是衰减的特征时间，常数设为：

$$p(t) = \begin{cases} 1 & \text{若 } t = 0 \\ IL(t) - TIL(t-1) & \text{若 } t > 0 \end{cases}$$

<sup>3</sup> TIL算法仅在如附件4所述、主要为静态图像内容的有限条件下进行了验证。在对更广泛的动态图像进行测试后，可能需要对TIL算法进行修订。

$$\tau = \begin{cases} 22 \frac{f}{24} & \text{若 } p(t) \geq 0 \\ 800 \frac{f}{24} & \text{若 } p(t) < 0 \end{cases}$$

其中：

$f$ ：视频的帧率（以Hz为单位）。

在附件3中显示了对TIL函数形式的推导过程，在附件4中描述了对衰减时间值的选择。

### 3 图像等级响应算法

图像等级响应（ILR）用于模拟人类视觉系统对图像流明度的响应情况。这对于评估由图像等级变化（例如，在换镜头时）引起的观看者烦恼的瞬时等级可能是有用的。它没有模拟亮度变化后主观烦恼的衰减。然而，它可以为计量设备提供有用的输入，当与本文件中规定的其他测量方法结合使用时，可模拟该衰减。它取决于图像的流明度和眼睛的适应等级。在ITU-R BT.2100建议书参考环境中，对该响应可以使用IL和TIL量来近似计算，如下所示：

$$ILR = \frac{(2^{IL})^{n_c}}{(2^{IL})^{n_c} + (2^{TIL})^{n_c}}$$

其中：

$ILR$ ：人类视觉系统的近似响应；

$n_c$ ：常数等于0.57。

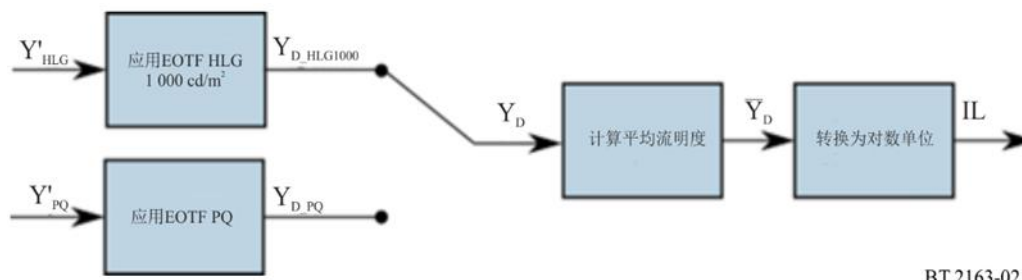
## 附件1 (资料性)

### 近似计算显示流明度以简化硬件实现

本附件描述了一种计算显示流明度分量 $Y_D$ 的近似方法，该方法简化了在第1段中所述的IL算法的实现。在硬件资源受限的情况下，该方法可在仅用于视觉辅助的设备中实现。使用这种近似方法，饱和色可能会导致较低的测量值。使用这种方法的设备将不适合对IL值进行数值分析，因此将不要求取IL值的对数。

图2显示了近似算法的框图。其输入是非线性HLG或PQ亮度信号 $Y'$ ，范围为[0:1]。如果信号是 $IC_7C_P$ ，则可以使用 $I$ 信号来代替下图中的 $Y'$ 。

图2  
显示近似IL测量算法的框图



BT.2163-02

### 转换为近似4线性显示光

对于HLG信号 $Y'_{HLG}$ ，ITU-R BT.2100建议书表5中规定的HLG EOTF直接应用于 $Y'_{HLG}$ ，对于标称峰值流明度为1 000 cd/m<sup>2</sup>的显示器，使用ITU-R BT.2100建议书表5注释5f中规定的 $\gamma$ 值1.2。所得到的信号近似于在[0:1 000] cd/m<sup>2</sup>范围内的显示线性光 $Y_{D\_HLG}$ 。

对于PQ信号 $Y'_{PQ}$ ，ITU-R BT.2100建议书表4中规定的PQ EOTF直接应用于 $Y'_{PQ}$ 。所得到的信号近似于在[0:10 000] cd/m<sup>2</sup>范围内的显示线性光 $Y_{D\_PQ}$ 。

近似线性显示光信号被称为 $Y_D$ ，而不管该信号是源自HLG还是PQ。

$Y_D$ 的近似值随后用于第1.2和1.3段中所述的其余计算。

## 附件2 (资料性)

### 用于高动态范围电视中亮度测量的客观指标

本附件描述了背景研究，该研究为开发第1段中所述的图像等级算法提供了信息。在对相关工作进行概述后，描述了一种主观测试，该测试建立了用作一组测试图像的亮度数据的值。然后，这些值用于测试许多潜在的客观指标，以测量图像亮度。

#### 相关工作

有大量关于亮度感知和适应的工作，这在BBC白皮书341 [1]中有所描述。

#### 为测试图像建立亮度值的实验

亮度一词表示“一个区域显示光的程度” ([2], 第69页)。它不同于明度，明度与物体的表面反射率有关，与物体的照明方式无关 ([2], 第70页)。亮度是一个无法直接测量的主观量，因此开发了一种主观测试方法来创建一组亮度基础真值。目的是基于显示的流明度值，创建一个与主观结果尽可能匹配的客观亮度指标。

<sup>4</sup> 注意：为了准确计算显示流明度，ITU-R BT.2100建议书中的EOTF应用于 $R'G'B'$ 信号分量，而不是此处所述的流明度分量。

要求测试对象调节蓝灰色的亮度，直到它与测试图像的感知整体亮度相匹配。蓝灰色的流明度是已知的，并可用作代表图像亮度的数值。受试者能够在测试图像与蓝灰色之间自由切换，并能够根据需要花费尽可能多的时间。

测试图像与蓝灰色显示在SIM2 HDR47E显示器上，使用其校准的LogLUV模式。蓝灰色等级范围为0至4 000 cd/m<sup>2</sup>，400个步长遵循指数函数，因此，步长在黑色时为 $3.9 \times 10^{-10}$  cd/m<sup>2</sup>，在顶端时为50 cd/m<sup>2</sup>。与所有LCD屏幕一样，当以单一颜色显示全屏时，SIM2无法准确显示输入亮度，尤其在高流明度等级下。以10 cd/m<sup>2</sup>的间隔记录输入蓝灰色等级的屏幕实际流明度，并在呈现结果之前将预期的蓝灰色值映射到这些测量值（必要时进行插值）。两个可调LED灯照亮显示器后的墙壁，使得从墙壁反射的光在5 cd/m<sup>2</sup>下测得D65白色。灯置于屏幕后，朝向墙壁，以尽量减少直接落在屏幕上的光。房间里没有其他光源。测试设置如图3所示。

图3

测试室设置





图4  
用于测试的图像



图4的注释：图像1至12<sup>5</sup>来自[3]，图像13来自[4]<sup>6</sup>，图像14和15由BBC研发部门创建。

SIM2显示器的LogLUV输入没有亮度控制，因此，普通的PLUGE信号不能用于校准黑色等级。因此，使用一组专门生成的测试信号（包括黑色等级范围内的亚黑和超黑）来找寻所需的偏移，并在显示前将该偏移添加到测试图像中。发现黑色等级偏移为 $0.005 \text{ cd/m}^2$ 。如果灯置于屏幕前而非屏幕后，则这比预期的要低。

<sup>5</sup> 这些图像经Mark Fairchild许可转载于此。

<sup>6</sup> 图像经斯图加特传媒大学许可转载于此。

在本研究中使用了15幅图像，如图4所示。前12幅图像来自Mark Fairchild的《HDR摄影调查》[3]，还有1幅来自斯图加特传媒大学的图像（第13幅）[4]以及和2幅由BBC研发部门制作的图像（第14和15幅）作为补充。由于原始图像的动态范围大于HDR电视中所用的预期范围，因此对图像进行了缩放，减小动态范围，使之看起来更具美感（由少数专业观看者判断）。这相当于调节相机光圈。测试集包括亮的图像和暗的图像，以及若干既有光亮区域又有阴影区域的图像。这些图像被转换为ITU-R BT.2100/BT.2020建议书的原色，并以1 920×1 080像素的分辨率进行显示，以匹配显示器的最大分辨率。

每幅图像显示在四个峰值显示流明度等级上，分别为500、1 000、2 000和4 000  $\text{cd/m}^2$ ，采用参考场景的流明度缩放方法。根据ITU-R BT.2100建议书表5注释5f，应用了适合显示峰值流明度的伽马函数。这只是增加测试中所用亮度范围的一种方法，并不限制结果对参考场景的系统的适用性。

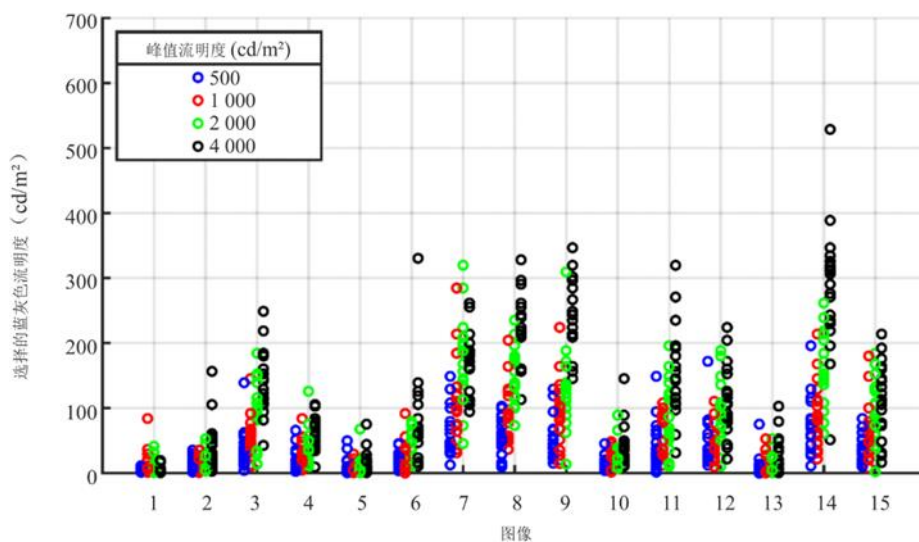
受试者坐在距显示器1.9米远的地方，这相当于屏幕高度的3.2倍。每个受试者在测试前都接受了正常视力筛查，然后给出了书面说明。测试开始时提供2幅训练图像和3幅“虚拟”图像。抛弃训练图像和虚拟图像的结果。给每个受试者的图像以不同的随机顺序呈现，并且注意同一幅图像（不同亮度）不会连续出现两次。20名受试者完成了测试。

### 为测试图像建立亮度值的实验结果

图5显示了所有20名受试者在每个峰值显示流明度下对每幅图像的单独响应情况。应当注意的是，即使在不同的峰值显示流明度等级下，图像也可能不都具有不同的主观亮度等级。例如，测试图像1（参见图4）主要是暗的，有一些强烈的高光。当显示流明度增加时，最大的主观差异在对高光的感知亮度上，因此，如果一小块高光区域的等级在感知上不显得重要，那么该图像的整体亮度可能不会受到显示峰值流明度变化的影响。这正是为了开发有效的亮度指标而需要研究的效果。尽管如此，结果显示，随着显示峰值流明度的增加，图像呈现更亮的总体趋势，并且显示，在测试集中亮的图像和暗的图像存在一定的扩张情况。

图5

根据图像和峰值显示流明度，为单个测试对象选择的蓝灰色流明度等级。  
每个圆圈对应一个受试者对该图像的评级。



## 客观指标

出于本研究的目的，开发了将显示的像素流明度值与从主观测试中收集的整体感知亮度值相关联的模型。如果将之用于信号监测，则该模型最终将需要对信号值而不是显示的光等级进行操作，但在此阶段，它与信号格式无关，因此，可将之应用于任何HDR图像。

测试图像存储为混合对数伽马 $Y'Cb'Cr'$ 图像，采用4:2:2颜色二次采样。在对色差分量进行上采样并转换为 $R'G'B'$ 之后，根据ITU-R BT.2100建议书表5计算显示流明度值。首先，移去HLG光电传递函数，以找寻场景线性光信号 $R_S G_S B_S$ ，然后根据显示峰值流明度应用伽马和缩放，以找寻显示颜色分量 $R_D G_D B_D$ 。最后，使用ITU-R BT.2100/BT.2020建议书颜色公式，从显示颜色分量计算得到显示流明度值。

以下模型从我们测试图像的显示像素流明度值中产生整体亮度的数值。对于特定像素，显示的流明度被定义为 $(Y_D(i, j))$ ，其中， $i$ 和 $j$ 为像素索引， $i \in 0:M-1$ 和 $j \in 0:N-1$ 。对于我们的图像， $M = 1080$ 和 $N = 1920$ 。

### 1) 平均显示流明度

$\frac{1}{MN} \sum_i \sum_j Y_D(i, j)$ 。基线指标，计算所有显示的像素流明度值的平均值。

### 2) 平均log10显示流明度

$\frac{1}{MN} \sum_i \sum_j \log_{10}(Y_D(i, j))$ 。继Fechner之后[5]。

### 3) 显示流明度的平均PQ逆EOTF

$\frac{1}{MN} \sum_i \sum_j \text{EOTF}^{-1}(Y_D(i, j))$ ， $\text{EOTF}^{-1}$ 如ITU-R BT.2100建议书表4中所定义。

### 4) 平均显示流明度提高到某个幂

$\frac{1}{MN} \sum_i \sum_j (Y_D(i, j))^p$ 。继Stevens之后[6]。在此，从 $p = 0.2$ 到 $p = 1$ 对值进行测试。

### 5) 平均CIE 1976明度

$\frac{1}{MN} \sum_i \sum_j L^*(Y_D(i, j))$ ， $L^*$ 如CIE所定义[7]，包括低流明度等级下的线性部分。使用对应作为参考白色的、75%信号等级的显示光值，因为这已被定义为混合对数伽马HDR的图形参考等级。对500、1000、2000和4000 cd/m<sup>2</sup>的峰值亮度等级，这些等级分别为120、203、344和581 cd/m<sup>2</sup>。

### 6) 加权平均显示流明度

$\frac{1}{MN} \sum_i \sum_j Y_D(i, j) \cdot \cos \theta_{ij} / \theta_{ij}^2$ ，其中， $\theta_{ij}$ 是像素 $(i, j)$ 与屏幕中心之间的眼对角，最小值为0.75°。该指标遵循Moon和Spencer [8]。在此处使用的形式中，假设观看者注视的是屏幕中心，并近似认为所有像素对着视网膜上的相同角度，省略了不会影响相关性系数的常数。

### 7) 屏幕中央的平均值

$\frac{4}{MN} \sum_{i=M/4}^{3M/4} \sum_{j=N/4}^{3N/4} Y_D(i, j)$ 。Moon和Spencer加权的简化版：仅计算屏幕中央四分之一那些像素的平均流明度。

### 8) 百分位数

第 $n$ 个百分位数 $P_n$ 指的是所有像素流明度等级中 $n\%$ 低于的流明度等级。这些与显示流明度等级的分布有关。测试从 $P_{10}$ 到 $P_{100}$ 的百分位数。

## 9) 百分位数范围

计算四分位范围 $P_{75} - P_{25}$ 以及第90个百分位数与第10个百分位数之间的差值 $P_{90} - P_{10}$ 。这些描述了显示流明度值的扩张情况。

## 10) 规定范围内的平均值

仅计算 $P_{25}$ 与 $P_{75}$ 之间那些值的平均显示流明度。还测试从 $P_{10}$ 到 $P_{90}$ 的范围。

60幅测试图像（在4个峰值亮度上有15幅图像，参见图4）的平均选择的灰色等级被用作用于评价模型的亮度基础真值。报告了每个模型与基础真值之间的Pearson相关性系数和Spearman等级相关性系数。

**客观指标与主观结果的相关性**

所有亮度指标的Pearson相关性系数和Spearman等级相关性系数如表1所示。结果也在图6和图7中以图形的形式做了呈现。

最简单的指标，平均显示流明度（指标1），是一个非常好的匹配，Pearson相关性系数略低于0.96。其性能如图8（左）所示。指标2-5中的各种非线性缩放方法仅导致数据偏离指标1的直线，如图8（右）中指数为0.33的指标5所示，其中，曲线显然更适合。唯一的例外是在求平均值之前将像素流明度值提高到0.8与0.9之间的某个幂（参见图7（左）），这比指标1略有改善。

依赖于观看者注视情况的平均加权显示流明度（指标6）表现不佳，屏幕中心处各值的平均值（指标7）也远不如所有值的平均值有效。这表明，靠近屏幕边缘的像素对整体亮度有重要作用。

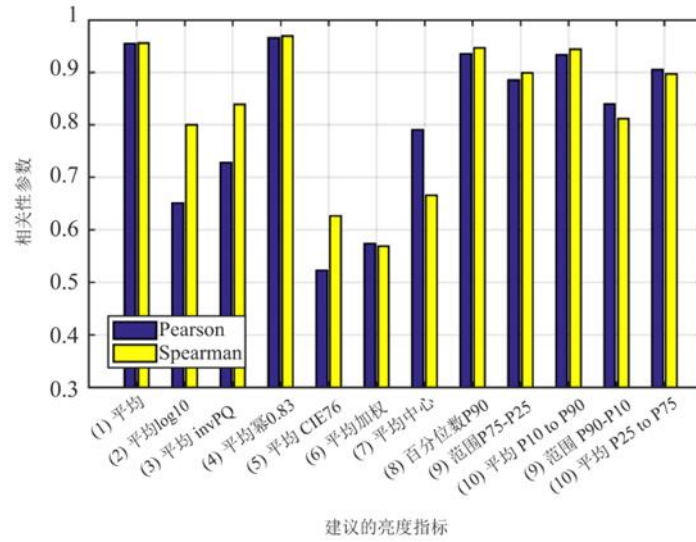
较低的百分位数（指标8）表现很差，但较高的百分位数与主观测试结果有很好的相关性。这表明，图像最暗部分的流明度分布不是感知整体亮度的重要因素，但最亮部分的等级是重要的驱动因素。第90个以上百分位数的相关性下降，这意味着，非常小的亮的区域对整体亮度的影响较小。

百分位数范围（指标9）的表现类似于对应于范围上限的百分位数，即 $P_{90} - P_{10}$ 的相关性类似于 $P_{90}$ 单独的相关性， $P_{75} - P_{25}$ 的相关性位于 $P_{70}$ 与 $P_{80}$ 之间。二者都与主观结果有相当好的相关性。我们的所有测试图像都包含一些暗的区域，因此，在这种情况下，百分位数范围并不比单个百分位数更具信息量。

规定范围内的平均值（指标10）也表现得比较好。范围越宽，相关性越高，即与指标1最相似的范围。

图6

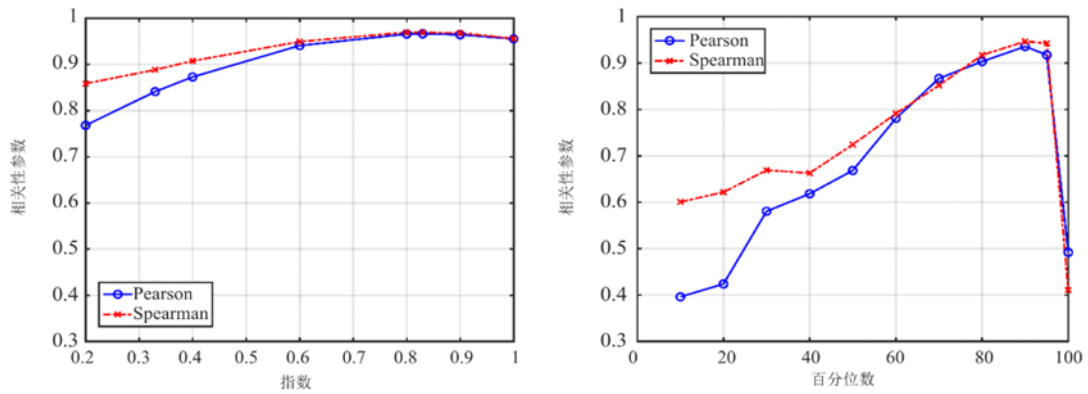
所有研究指标的相关性系数  
给出最高相关性的参数用于指标4和8



BT.2163-06

图7

指标4的相关性系数，平均显示像素流明度值提高到某个幂，指数从0.2到1（左）；  
指标8的相关性系数，显示像素流明度值的百分位数10到100（右）



BT.2163-07

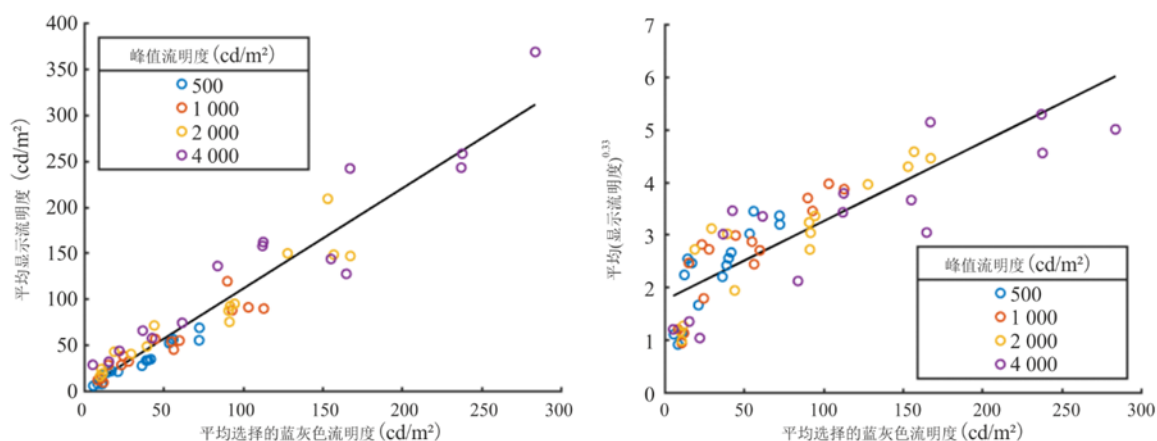
表1

所研究亮度指标的Pearson相关性系数和Spearman等级相关性系数。  
表中各部分中的最佳相关性以粗体形式突出显示

指标	参数值 (如果适用)	Pearson 相关性系数	Spearman等级 相关性系数
1) 平均显示流明度	-	<b>0.955378</b>	<b>0.955877</b>
2) 平均log10显示流明度	-	0.651608	0.800111
3) 显示流明度的平均PQ逆EOTF	-	0.728708	0.839344
4) 平均CIE 1976明度	-	0.523273	0.626341
5) 提高到某个幂的平均显示流明度	0.2	0.767964	0.858405
	0.33	0.841025	0.888358
	0.4	0.872807	0.907530
	0.6	0.940758	0.949264
	<b>0.8</b>	0.965687	<b>0.969492</b>
	<b>0.83</b>	<b>0.966131</b>	0.969214
6) 平均加权显示流明度	-	0.573798	0.568769
7) 屏幕中央的平均值	-	<b>0.791089</b>	<b>0.666296</b>
8) 百分位数	$P_{10}$	0.395950	0.600389
	$P_{20}$	0.423570	0.621506
	$P_{30}$	0.580292	0.668797
	$P_{40}$	0.617995	0.662573
	$P_{50}$	0.668203	0.724034
	$P_{60}$	0.780425	0.790886
	$P_{70}$	0.866544	0.851570
	$P_{80}$	0.903016	0.916921
	<b><math>P_{90}</math></b>	<b>0.935327</b>	<b>0.946374</b>
	$P_{95}$	0.917317	0.942317
9) 百分位数范围	$P_{75} - P_{25}$	0.885345	0.899194
	<b><math>P_{90} - P_{10}</math></b>	<b>0.934097</b>	<b>0.944540</b>
10) 规定范围内的平均值	$P_{25} - P_{75}$	0.839611	0.811892
	<b><math>P_{10} - P_{90}</math></b>	<b>0.905243</b>	<b>0.89686</b>

图8

指标1, 平均显示流明度 (左), 指标5, 平均显示流明度提高到0.33的幂 (右),  
针对每幅测试图像, 依据平均选择的蓝灰色流明度 (用作主观亮度的测量, 详见附件2) 进行绘制



BT.2163-08

## 结论

研究了将显示的像素流明度值与整体图像亮度相关联的十类客观亮度指标。一组HDR图像的感知整体亮度的先前测量结果确定了用于评价所研究指标的基础真值数据。

在计算平均值之前, 性能最好的指标将显示的像素流明度值提高到0.83的幂, 但这仅略好于直接找到平均显示像素流明度值的更简单的指标。首先将这些值提高到某个幂所带来的改进非常小, 并可能在实验不确定性的限制范围内, 因此实时应用更倾向于使用更简单的方法。0.96的高相关性表明, 这一简单的指标将成为亮度监测器的一个有效基础。

## 参考文献

- [1] Noland, K.C., Pindoria M. and Cotton, A., "Modelling Brightness Perception for High Dynamic Range Television" Research & Development White Paper, 2019, pp. 1, 2. Available: <http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP341.pdf>
- [2] Hunt, R. W. G., "The Reproduction of Colour in Photography, Printing and Television," 4th ed. Fountain Press, 1987.
- [3] Fairchild, Mark, "The HDR Photographic Survey," website, <http://markfairchild.org/HDR.html>
- [4] Fröhlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A., and Brendel, H., "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," Proc. SPIE 9023, Digital Photography X, 90230X, 7 March 2014. Images available online at <https://www.hdm-stuttgart.de/vmlab/hdm-hdr-2014/>
- [5] Fechner, G. T., "Elemente der Psychophysik," Breitkopf und Härtel, 1860.
- [6] Stevens, S. S., "To Honour Fechner and Repeal His Law," Science, vol. 133, no. 3446, pp. 80-86, 1961.
- [7] Commission Internationale de l'Eclairage, "CIE 1976 L\*u\*v\* colour space," 2014. [Online] Available: <https://cie.co.at/eilv/159>.

- [8] Moon, P. and Spencer, D. E., "The Visual Effect of Non-Uniform Surrounds," J. Optical Soc. America, vol. 25, no. 3, pp. 233-248, 1945.

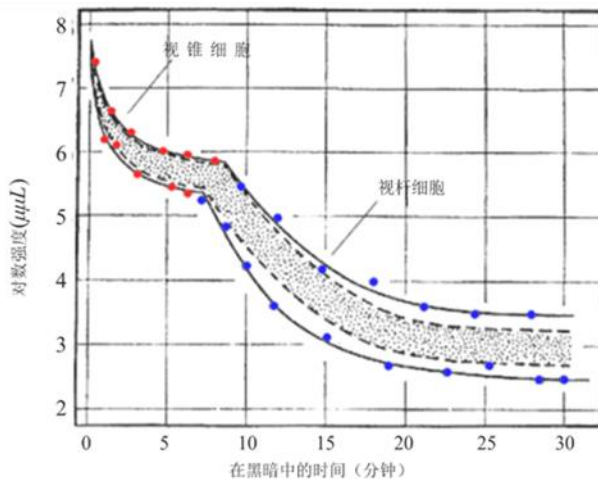
### 附件3 (资料性)

#### 时间图像等级函数形式的背景

##### 引言

人类视觉会随着时间的推移而适应所看到的内容。这种适应在某种意义上取决于观察者最近观看的内容。众所周知，对于以固定流明度和固定持续时间呈现的刺激，观察者的适应等级与呈现的流明度及其持续时间的乘积（即暴露于观察者的总能量）有关[1] [2] [3]。如果在完全适应这样的固定流明度等级后，移去刺激，则暗适应随后需要大约30分钟才能完全生效。图9显示了作为时间函数的暗适应曲线。

图9<sup>7</sup>  
暗适应曲线



BT.2163-09

图9的注释：阴影区域代表该组受试者的80%。（Hecht和Mandelbaum的数据来自Pirenne M. H., 《暗适应和夜视》。第5章。载于：Davson, H. (ed), The Eye, vol 2. London, Academic Press, 1962.）

可以看出，视杆细胞和视锥细胞适应相似的曲线，但处于不同的光状态。在视网膜中央凹中只存在视锥细胞，因此由视杆细胞决定的那部分曲线将会消失。如上所述，暗适应曲线取决于预适应流明度。此外，预适应流明度的持续时间对暗适应有影响。预适应流明度的持续时间越短，适应越快。这表明，越早暴露于流明度，对当前适应状态的影响越小。

<sup>7</sup> 该图取自：<https://webvision.med.utah.edu/book/part-viii-gabac-receptors/light-and-dark-adaptation/>.



可以假设，通过以加权方式对过去视频帧的流明度进行积分，可以近似得到暴露于视频内容的观察者的当前适应状态，这样，越早显示的帧，被赋予的权重越低。就图像处理而言，等效方法将是在一定数量的先前帧上单独积分每个像素位置。然而，这种积分相当于对每个像素位置应用时间低通滤波器。因此，原则上，这有可能通过将低通滤波器与视频本身卷积来确定暴露于视频的观察者的视觉系统的适应状态。

这种卷积在计算上是昂贵的，此外，可以以更有效和生物学上似乎更合理的方式来计算效果。为此，对（人类）大脑中神经元的响应可以通过（广义）泄漏积分发射（LIF - Leaky Integrate-and-Fire）模型来很好地建模。神经元显示输入阶段的神经元膜电流与输出阶段的膜电压之间的关系<sup>8</sup>。众所周知，神经元根据其膜电阻泄漏一些电势，因此，在时间 $t$ 上，驱动电流 $I(t)$ 与膜电压 $V_m$ 的关系如下所示：

$$I(t) = \frac{V_m(t)}{R_m} + C_m \frac{dV_m(t)}{dt}$$

其中：

$R_m$ ：膜电阻；  
 $C_m$ ：神经元电容。

这本质上是一个泄漏积分器<sup>9</sup>。可以乘以 $R_m$ ，并引入膜时间常数 $\tau_m = R_m C_m$ ，得出<sup>10</sup>：

$$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + R_m I(t)$$

合理的假设是，在时间 $t = 0$ 时，膜电压处于某个恒定值，即 $V_m(0) = V$ ，并且在此之后的任何时间输入都为零，即 $I(t) = 0$  对于  $t > 0$ 。这相当于神经元开始适应没有输入的情况。例如，对于光感受器来说，这将是暗适应开始时的情况，但要注意的是，该过程并不是光感受器所独有的。则上述公式的封闭解为：

$$V_m(t) = V e^{-\frac{t}{\tau_m}} \quad \text{对于 } t > 0$$

可以看出，该公式确实定性地模拟了图9的暗适应曲线。还要注意的，该公式基本上相当于Crawford在1947年提出的模型[4] [5]。泄露积分已被证明是人类视觉中涉及的神经元适应性行为的一个合适模型。

对于接近0的 $t$ 值，上述函数的导数趋于 $-\frac{V}{\tau_m}$ ，因此，可以通过参数 $\tau_m$ 来控制初始变化率。为了获得计算效率高的公式，对上述微分公式改写如下：

$$\tau_m (V_m(t) - V_m(t-1)) = -V_m(t) + R_m I(t)$$

故：

$$(\tau_m + 1)V_m(t) - \tau_m V_m(t-1) = R_m I(t)$$

从而导致：

$$V_m(t) = \frac{\tau_m}{\tau_m + 1} \left( V_m(t-1) + \frac{I(t)}{C_m} \right)$$

<sup>8</sup> [https://en.wikipedia.org/wiki/Biological\\_neuron\\_model#Leaky\\_integrate-and-fire](https://en.wikipedia.org/wiki/Biological_neuron_model#Leaky_integrate-and-fire).

<sup>9</sup> [https://en.wikipedia.org/wiki/Leaky\\_integrator](https://en.wikipedia.org/wiki/Leaky_integrator).

<sup>10</sup> <https://neurondynamics.epfl.ch/online/Ch1.S3.html>.

该公式的结构表明，神经元在时间 $t$ 的输出 $V_m$ 是神经元在时间 $t - 1$ 的输出以及在时间 $t$ 的输入 $I$ 的一个函数。

为了将该模型实现为可以应用于图像或从图像导出的值的泄漏积分器，可将膜电阻 $R_m$ 设为1，使得：

$$V_m(t) = \frac{\tau_m}{\tau_m + 1} \left( V_m(t - 1) + \frac{I(t)}{\tau_m} \right)$$

为了在广播设置中应用该模型，每帧的单个适应等级比每像素的适应等级更可取。这可以通过注意到帧 $t$ 的稳态适应 $L_a(t)$ 可由帧的某个平均流明度（例如，图像等级（ $IL$ ））来近似来实现。适应的时间状态 $L_T(t)$ 由以下公式给出：

$$L_T(t) = \frac{\tau_m}{\tau_m + 1} \left( L_T(t - 1) + \frac{L_a(t)}{\tau_m} \right)$$

应用这种方法的效果是时间低通滤波器的效果，尽管没有与这种滤波器操作相关联的计算复杂性。因此，它被作为人类视觉适应移动内容的一种度量。

上述公式可重新排列为：

$$L_T(t) = L_T(t - 1) \left( 1 - \frac{1}{\tau_m + 1} \right) + L_a(t) \left( \frac{1}{\tau_m + 1} \right)$$

## 参考文献

- [1] F. A. Mote and A. J. Riopelle. The effect of varying the intensity and the duration of preexposure upon foveal dark adaptation in the human eye. *J. comp. physiol. Psychol.*, pp. 46(1):49-55, 1953.
- [2] Ledda, Patrick, Luis Paulo Santos, and Alan Chalmers. "A local model of eye adaptation for high dynamic range images." In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pp. 151-160. ACM, 2004.
- [3] Mote, F. A., and A. J. Riopelle. "The effect of varying the intensity and the duration of preexposure upon foveal dark adaptation in the human eye." *The Journal of general physiology* 34, No. 5 (1951): pp. 657-674.
- [4] Crawford, B. H. "Visual adaptation in relation to brief conditioning stimuli". *Proc. R. Soc. Lond. B* 134, No. 875 (1947): pp. 283-302.
- [5] Pianta, Michael J., and Michael Kalloniatis. "Characterisation of dark adaptation in human cone pathways: an application of the equivalent background hypothesis". *The Journal of physiology* 528, No. 3 (2000): pp. 591-608.

## 附件4 (资料性)

### 时间图像等级中时间常数的验证

#### 引言

本附件描述了为调查第2段中所载之衰减时间算法的适当 $\tau$ 而进行的研究。

#### 主观测试

杜比实验室开发了一套主观测试。该实验的新颖之处在于研究对亮度随时间变化的主观容忍度。开展了两个版本的实验来估算从亮-到-暗-到-亮以及从暗-到-亮-到-暗过渡的视觉响应情况。前者首先向观察者展示10秒钟的“亮”图像，然后切换到时长可变的“暗”图像，最后切换回“亮”图像7秒钟。以类似的方式，后者首先向观察者展示10秒钟的“暗”图像，然后切换到时长可变的“亮”图像，最后切换回“暗”图像7秒钟。表2显示了实验的两个版本中使用的中间图像的不同持续时间。在实验之前，进行了一系列Beta测试来确定用于各版本的最佳持续时间。

表2

实验中实现的适应时间过程的持续时间

实验版本	中间图像的持续时间 秒
亮-到-暗-到-亮	0.5, 5, 10, 25, 40, 60, 90, 120, 150
暗-到-亮-到-暗	0.5, 2, 5, 10, 15, 20, 30, 60

在兴趣过渡（从中间图像到最终图像）后，要求受试者对其体验进行评级。表3显示了与损伤等级相关的分值和术语。表示疼痛/烦恼程度的术语各有两个值来增加等级。此外，两个版本的测试使用不同的语言来反映观察者在亮度交界处后的感受。

表3

用于对两个实验版本中交界处进行评级的损伤等级

分值	术语 亮-到-暗-到-亮	术语 暗-到-亮-到-暗
1, 1.5	疼痛	烦恼
2, 2.5	显著的疼痛	显著的烦恼
3, 3.5	轻微的疼痛	轻微的烦恼
4	不疼痛	不烦恼

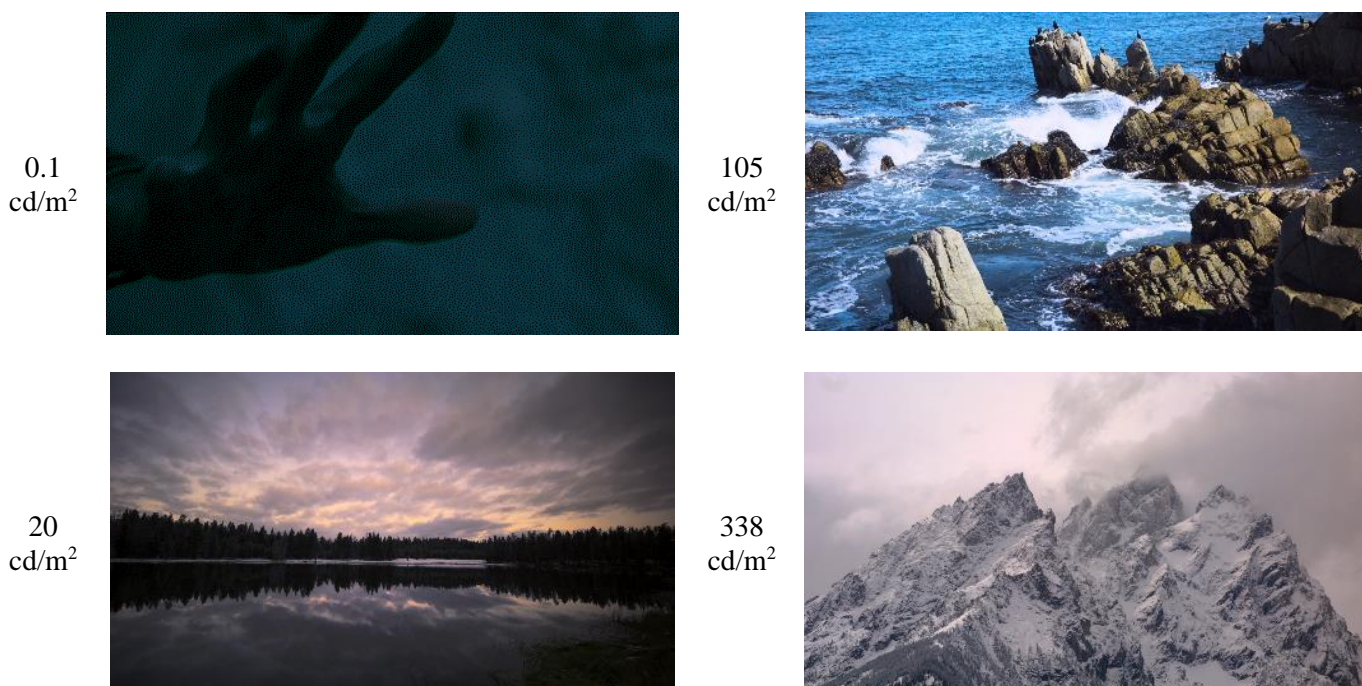
本实验的主观参与者由17名专家和非专家观看者混合组成。Christie 4K 6P激光投影仪系统采用了5 cd/m<sup>2</sup>环绕的ITU-R BT.2100建议书观看条件。激光投影仪校准到1 000 cd/m<sup>2</sup>的峰值流明度设置上，并具有0.0004 cd/m<sup>2</sup>的测量黑色等级。图像的位置与观察者的观察位置成3.2个图像高度的水平角。图像信号经PQ光电传递函数编码后，通过标准数字接口（SDI）发送。

## 测试图像

在这项研究中总共使用了四幅图像（如图10所示）。一幅图像作为“暗”图像，而另外三幅图像作为不同等级的“亮”图像。对每幅“亮”图像重复测试过程。对所有图像都进行了色调映射，以便在投影仪的流明度范围内显示。“暗”图像由杜比实验室公司（Dolby）制作，三幅“亮”图像由Spears & Munsil授权给Dolby<sup>11</sup>。图1显示了这些图像及其计算得到的平均显示流明度值。使用第1段中的图像等级算法来计算平均显示流明度值。

图10

在1 000 cd/m<sup>2</sup>投影仪上以规定的平均显示流明度值测试图像



## 测试结果

在实验的两个版本中，对三个不同的“亮”图像进行了评价，以测试向不同的平均显示流明度值的过渡情况。这样做是为了显示平均主观分值（MSS）如何随着亮度跳变幅度而变化，以及它如何随着适应的时间过程而变化。

图11显示了九次适应时间试验中三幅图像的平均评级和95%置信区间。MSS为4表示不会引起任何眼睛疼痛或主观烦恼的过渡，而2或以下分值则表示显著的眼睛疼痛或烦恼。

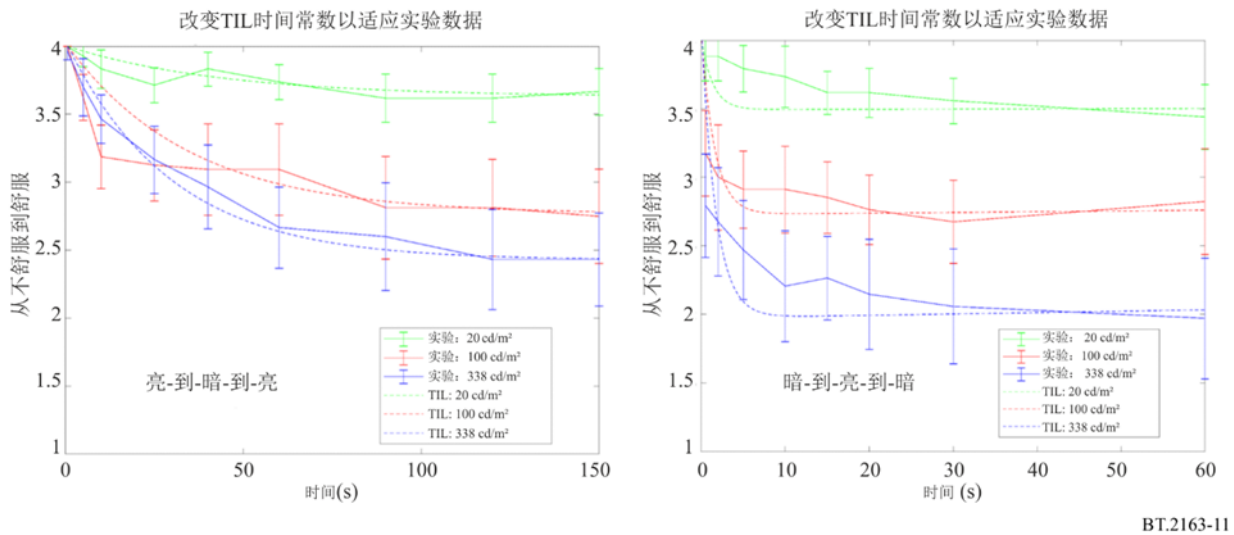
<sup>11</sup> Spears & Munsil超高清基准（2023年）。<https://www.biaslighting.com/products/spears-munsil-ultra-hd-benchmark-2023>

为了为TIL公式拟合一个适当的 $\tau$ ，需要将适应时间与MSS结果相关联。第3段中的ILR更接近主观体验。

因此，在从平均显示流明度转换为图像等级后，对TIL公式应用各种不同的 $\tau$ 来估算在最初十秒钟暴露于“亮”图像和不同适应时间后对“暗”图像的适应等级。一旦对每个“亮”图像和适应时间的组合做了TIL估算，就可以计算得到ILR值。然后对响应值进行评级，使其反映各自亮度值跨越的MSS范围。这通过标准范围归一化和评级程序来完成。

在图11中用虚线标出了这些TIL估算的MSS值和最佳 $\tau$ ，并用实线标出了实验数据。

图11  
相比近似生成的平均主观分值，不同适应时间的亮度过渡的平均主观分值



BT.2163-11

该函数形式的最佳 $\tau$ 值如下所示（在置信区间内）。为了保持较低的适应时间，对第2段，使用最佳值的下限。

$$\tau = \begin{cases} [22 \text{ to } 25] \frac{f}{24} & \text{若 } p(t) \geq 0 \\ [800 \text{ to } 2000] \frac{f}{24} & \text{若 } p(t) < 0 \end{cases}$$

时间衰减更接近于“亮-到-暗-到-亮”版本中实验数据的趋势。10秒时对100 cd/m<sup>2</sup>存在明显偏差。未来调查这种异常值的工作将会有所帮助。衰减时间常数不像在“亮-到-暗-到-亮”版本中那样清晰地遵循“暗-到-亮-到-暗”版本中的实验数据。这很大程度上是源自亮度交界处幅度在短时间内带来的影响增加。即使最短持续时间的“亮”图像，观察者也会立即注意到一个后像。当前的函数形式不允许立即下降然后缓慢衰减。进一步研究这一现象将会有所帮助。