

国 际 电 信 联 盟

**ITU-R**

国际电联无线电通信部门

**ITU-R BT.2021 建议书**

(08/2012)

**立体三维电视系统的  
主观评价方法**

**BT 系列  
广播业务  
(电视)**



ITU 国际电信联盟

## 前言

无线电通信部门的职责是确保卫星业务等所有无线电通信业务合理、平等、有效、经济地使用无线电频谱，不受频率范围限制地开展研究并在此基础上通过建议书。

无线电通信部门的规则和政策职能由世界或区域无线电通信大会以及无线电通信全会在研究组的支持下履行。

## 知识产权政策 (IPR)

ITU-R的IPR政策述于ITU-R第1号决议的附件1中所参引的《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/en>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 系列建议书

(也可在线查询 <http://www.itu.int/publ/R-REC/en>)

系列	标题
<b>BO</b>	卫星传送
<b>BR</b>	用于制作、存档和播出的录制；电视电影
<b>BS</b>	广播业务（声音）
<b>BT</b>	<b>广播业务（电视）</b>
<b>F</b>	固定业务
<b>M</b>	移动、无线电定位、业余和相关卫星业务
<b>P</b>	无线电波传播
<b>RA</b>	射电天文
<b>RS</b>	遥感系统
<b>S</b>	卫星固定业务
<b>SA</b>	空间应用和气象
<b>SF</b>	卫星固定业务和固定业务系统间的频率共用和协调
<b>SM</b>	频谱管理
<b>SNG</b>	卫星新闻采集
<b>TF</b>	时间信号和频率标准发射
<b>V</b>	词汇和相关问题

**说明：** 该ITU-R建议书的英文版本根据ITU-R第1号决议详述的程序予以批准。

电子出版  
2013年，日内瓦

© 国际电联 2013

版权所有。未经国际电联书面许可，不得以任何手段复制本出版物的任何部分。

## ITU-R BT.2021 建议书

## 立体三维电视系统的主观评价方法

(2012年)

**范围**

本建议书阐述了立体三维电视（3DTV）系统的评价方法，其中包括一般测试方法、分级量表和观看条件。

国际电联无线电通信全会，

考虑到

- a) 目前已收集了大量有关各实验室用于评估3DTV系统关键性能特性的各种方法的信息；
- b) 对这些方法的研究表明，不同的实验室之间在测试的诸多方面存在着相当程度的一致性；
- c) 采用一种标准的方法，对各实验室之间的信息交换极为重要；
- d) 引入3DTV业务可能需要开发新的图像格式、图像处理和传输技术，而这些内容的性能均需要通过主观方法进行评价，

建议

**1** 应将下文附件1所述有关立体3DTV图像质量评估的一般测试方法、分级量表和观看条件用于实验室的实验，并在可能的情况下用于操作评估。

**附件1****1 评估（感知）的角度**

立体3DTV通过重现视觉场景下感知物体相对深度的条件，充分利用了人类双眼视觉系统的特性。当前立体成像的主要要求是能用两部水平排列的摄像机捕获同一场景的至少两张视图。场景中描绘的物体图像在左右视图中将有不同的相对位置。两个视图中相对位置的差异通常被称为图像像差（或视差），一般用像素、物理距离（例如，毫米）或相对测量值（例如，与屏幕宽度的百分比）来表示。应将图像像差与角度（视网膜）像差区分开来。事实上，相同的图像像差信息，会因观看距离不同，而产生不同的角度（视网膜）像差。深度感知的程度与方向基于立体图像产生的视网膜像差的程度和方向。

一般，适用于单视场电视图像的评估要素，如分辨率、彩色再现、运动表现、整体质量、清晰度等亦可适用于立体电视系统。此外，立体电视系统还有许多特有的评估要素，这可包括深度分辨率（深度方向上的空间分辨率）、深度运动（沿深度方向的运动或移动是否顺利再现）以及空间畸变等要素。对于后一种要素，木偶剧效应（*puppet theatre effect*）（即感知的物体异乎寻常地大或小）和纸板效应（*cardboard effect*）（即感知的物体是立体的，但却异乎寻常地扁）是两个众所周知的例子。

可以确定三个共同影响立体系统所提供的体验质量的基本感知角度：图像质量、深度质量和视觉舒适度。一些研究人员认为，亦可从更普遍的概念（如自然度和临场感）角度衡量立体成像技术的心理影响。

### 主要感知角度

图像质量是指系统所提供图像的感知质量。这是视频系统性能的一个主要决定因素。图像质量主要受技术参数以及编码和/或传输过程等产生的误差的影响。

深度质量是指系统提供增强的深度感的能力。即使是在标准2D图像中，单眼视觉线索（如线性透视、模糊、梯度等）也可传达一定的深度感。但立体3D图像亦包含像差信息，提供额外的深度信息，因此比2D的深度感更强。

视觉（不）舒适度是指对观看立体图像相关（不）舒适度的主观感觉。拍摄不当或显示不当的立体图像可造成严重的不舒适感。

### 额外感知角度

自然度是指感知立体图像是现实的如实再现（即感知现实主义）。立体图像可能会出现各种不同的畸变，使其看起来不太自然。例如，有时立体物体看起来异乎寻常地大或小（木偶剧效应），或看起来异乎寻常地扁（纸板效应）。

临场感是指仿佛身临另一个地方或环境的主观体验。

本建议书提供了关于评估上述三个主要角度（图像质量、深度质量和视觉舒适度）的方法和程序的信息。自然度和临场感的评估方法未包含在本建议书中，但计划稍后会将这些方法纳入建议书。

## 2 主观方法

ITU-R BT.500建议书列出了许多图像质量评估方法。所有方法都是进行一系列的评判试验，向一组观看者播放用被研究系统（如使用不同参数的算法；使用不同比特率的编码技术；不同的传输情形等）处理过的视频序列集。在每次试验中，观看者被要求用规定的量表评估视频序列的特性（如图像质量）。这些方法各不相同，主要是演示模式（即向观看者播放视频序列的方式）和观看者用来对这些序列进行评分的量表不同。



测试图像是根据第4段所述项目选择的双目立体图像。评价者对下列三项进行评估：

- 图像质量：对测试图像进行处理的系统以及用于显示拟评估图像的显示器对立体3D图像分辨率的影响；
- 深度质量：对测试图像进行处理的系统以及用于显示拟评估图像的显示器对立体3D图像深度感知的影响；
- 视觉舒适度：对测试图像进行处理的系统以及用于显示拟评估图像的显示器对立体3D图像观看舒适度的影响；

本建议书包括ITU-R BT.500建议书中的四种方法，其中三种方法已在过去二十年中成功用于解决与立体成像技术的图像质量、深度质量和视觉舒适度相关的各种研究问题。这些方法是：

- 单刺激（SS）法
- 双刺激连续质量量表（DSCQS）法
- 刺激比较法
- 单刺激连续质量评价（SSCQE）法

在使用这些方法时酌情对其进行了细微修改，如对视觉舒适度使用了不同的量表。与图像质量、深度质量和视觉舒适度的评价方法相关的演示模式和量表分别见表1、2、3。

本节中将对每种方法进行简要描述。所有方法中共同的方法要素在随后几节中介绍。

## 2.1 单刺激（SS）法

这一程序由一系列的评判试验组成，这些试验可酌情分成几个测试阶段，中间为休息时间。在每次试验中，只演示一个“测试”视频序列，即用一种被研究系统处理过的序列，并用规定的量表进行评分。

### 2.1.1 SS法的试验结构

在每次试验中，在演示拟评估的“测试”视频序列前后分别演示中灰场，前面的中灰场可包含一个固视目标，如试验编号，像差为零，持续时间应 $\leq 3$ 秒。后面的中灰场可包含评价提示语，如“马上评分”，且应持续足够时间，以便观看者完成打分（如 $\leq 10$ 秒）。“测试”视频序列一般应持续约10秒<sup>1</sup>。典型SS试验的结构见图1。

---

<sup>1</sup> 一些研究人员提倡使用持续时间较长的序列，主要依据是一种假设，即充分鉴赏立体内容需要的时间比鉴赏一般的平面（2D）内容长。迄今为止，并无多少经验证据能够证实或驳斥这种说法。

### 2.1.2 SS法的分级量表

对于图像质量评估，可使用两种有质量标示的量表：非连续五级量表和国际电联标准的连续质量量表（见表1）。质量标示为“优”、“良”、“中”、“差”、“劣”。

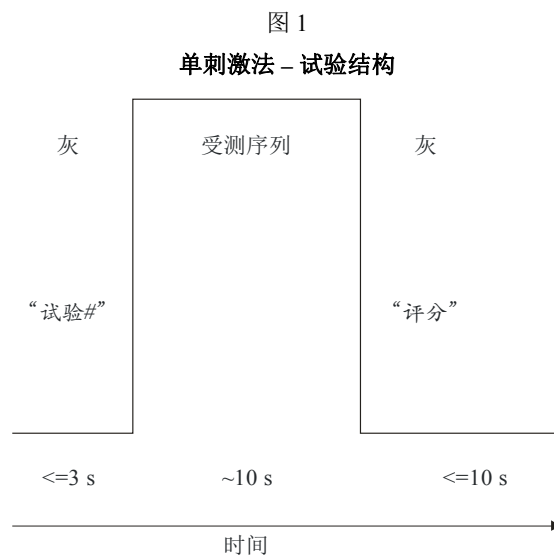
深度质量评估亦可使用相同的量表（见表2）。在这种情况下，观看者被要求评价深度表征的质量，而非图像本身的质量。

对于视觉舒适度的评估，可使用两种舒适度标示的量表：非连续五级量表和连续舒适度量表（见表3）。舒适度标示包括：“非常舒适”，“舒适”，“轻度不舒适”，“不舒适”和“非常不舒适”。

### 2.1.3 SS法的意见分数据

实验中每个序列的评分称为“意见分”。通常针对每个被研究系统计算这些分数的平均值，称为平均意见分（MOS）。

“参考”视频序列是测试序列未经任何处理的版本（见第8段），可包含在序列集中。将“参考”视频序列纳入序列集，就可计算“意见分差值”，即研究中每个序列的“测试”和“参考”版所得评分之间的算数差。针对每个被测试系统得到的意见分差值平均数称为平均意见分差值（DMOS）



BT.2021-01

## 2.2 双刺激连续质量量表（DSCQS）法

本程序由一系列的评判试验组成，这些试验可酌情分成几个测试阶段，中间为休息时间。在每次试验中，同一视频序列的两个版本各演示两次，总计四次演示。测试系统的一般性安排应如图2所示。

### 2.2.1 测试素材的演示

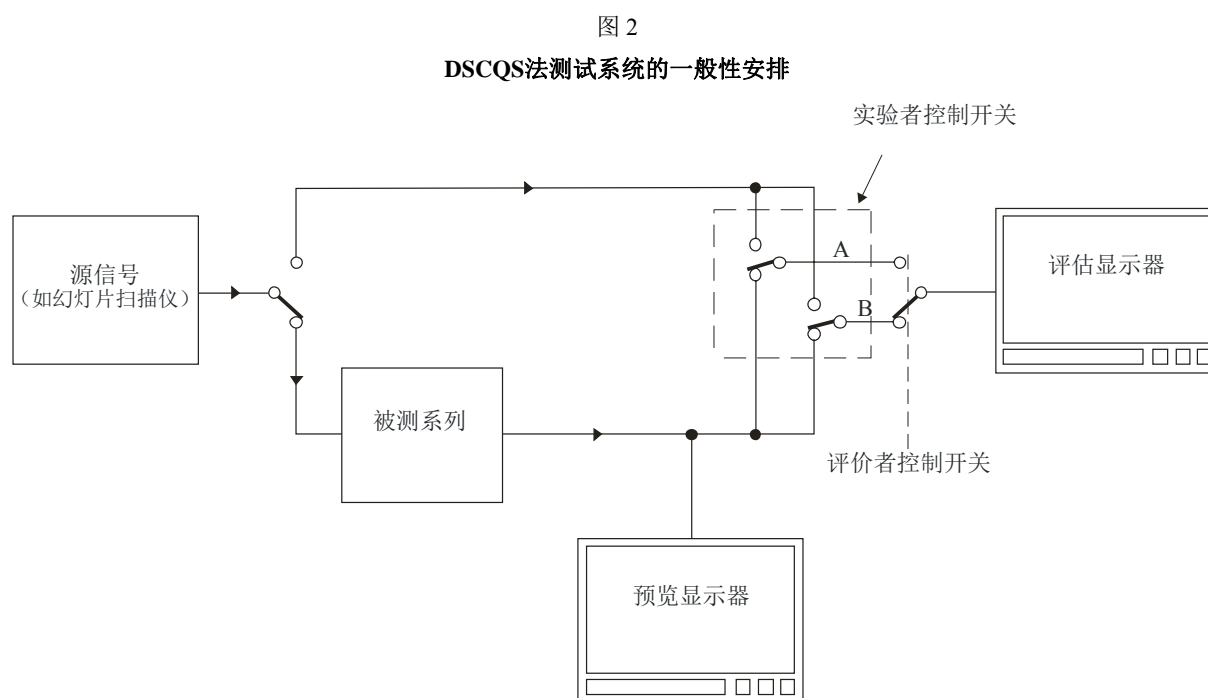
一个测试阶段由多次演示组成。对于只有一位观察者的演示结构类型I，每次演示时评价者可在A和B两种信号之间自由切换，直到评价者得出每个信号相关质量的心理尺度。通常评价者可以在10秒的时间内选择切换两或三次。对于同时有多位观察者的演示结构类型II，在记录结果之前，条件对要显示一次或多次，每次持续时间相同，以便让评价者得出这一对条件相关质量的心理尺度，然后再把条件对显示一次或多次，同时记录结果。重复的次数取决于测试序列的长度。对于静态图像，可能宜使用3-4秒的序列并重复5次（在最后2次期间评分）。对于受到时变伪影影响的动态图像，可能宜使用10秒的序列并重复2次（在第2次重复期间评分）。演示结构见图5。

### 2.2.2 DSCQS法的分级量表

在DSCQS法中，观看者被要求对A和B两个视频序列进行评分。对于图像质量和深度质量评估，可使用国际电联标准的连续质量量表（见表1和表2）。对于视觉舒适度评估，应使用舒适度标示为“非常舒适”、“舒适”、“轻度不舒适”、“不舒适”和“非常不舒适”的连续舒适度量表（见表3）。

### 2.2.3 DSCQS法的意见分数据

每次试验中得到的每个序列“测试”和“参考”版的评分用来计算意见分差值。再用后者计算每个被研究系统的DMOS。



BT.2021-02

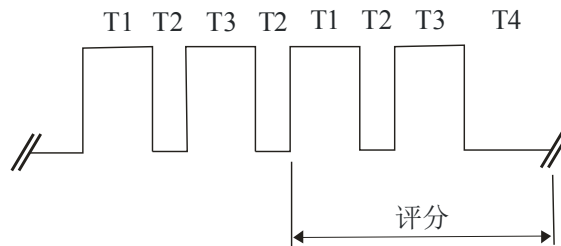
这种方法有下述I和II两种演示结构类型。

类型I：评价者一般是单独一个人，可以在A和B两种条件之间切换，直到对每一种条件的评分都感到满意为止。A线路和B线路都提供了直达参考图像，或通过被测系统提供图像。但哪条线路得到哪个图像在一个测试条件和下一个测试条件之间是随机变化的，它们由实验者注明，但不公布。

类型II：来自A线路和B线路的图像连续显示给评价者，供评价者对每一图像进行打分。对于每次演示，A线路和B线路都像上述类型I那样得到图像。此类型的质量范围有限，其结果的稳定性被认为仍在研究之中。

图 3

## 双刺激连续质量量表法 - 试验结构



演示阶段：

T1 = 10 s	测试序列 A
T2 = 3 s	中灰级
T3 = 10 s	测试序列 B
T4 = 5-11 s	中灰级

BT.2021-03

## 2.3 配对比较 (PC) 法

在PC法中，对一组“测试”序列，即经不同系统（如不同的比特率、不同的算法等）处理过的序列，进行配对比较（即两个一起）。要求观察者作出判断，在测试场景下这一对序列中哪个要素是首选。所需判断的数目是被研究系统数目的函数。通常被测系统（X、Y、Z等）按所有可能的 $n(n-1)$ 组合进行排列（XY、ZY、YZ等）。此外，所有序列对应以两种可能的顺序显示（如XY、YX）。

### 2.3.1 PC法的试验结构

试验首先演示一个中灰场，其中可能包含一个固视目标，如试验编号，像差为零，持续时间应 $\leq 3$ 秒。接下来演示拟比较的序列。每个受测序列一般应持续约10秒，可在两个显示器上同步演示（或同一显示器上并排演示）或在同一显示器上相继演示（如AB）。在后一种情况下，两个序列之间插播长3秒的中灰场。试验结束时显示中灰场，其中可包含要求评分提示语，如“马上评分”，且应持续足够长的时间，以便观看者完成打分（如 $\leq 10$ 秒）。典型PC试验示例见图4。

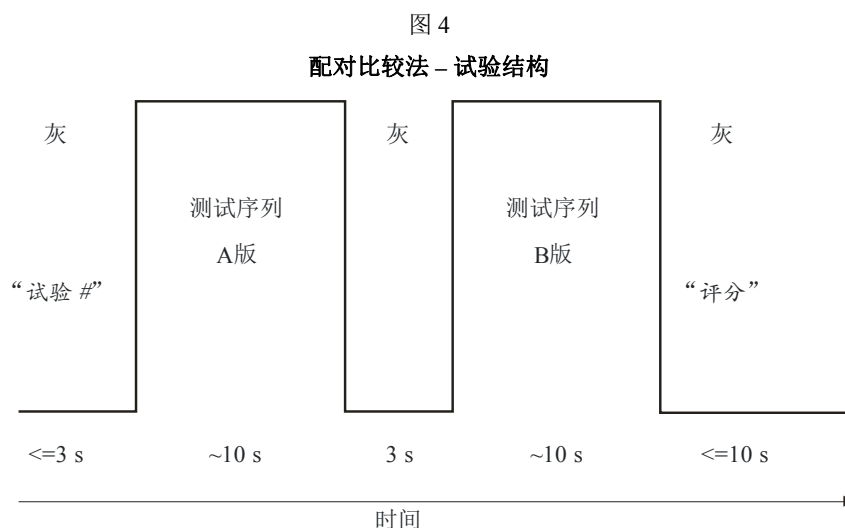


### 2.3.2 PC法的分级量表

观看者可能被要求用一个二进制量表（如A是首选）提供一个简单的偏好判断，或提供一个分级偏好（如A优于B）。图像质量、深度质量和视觉舒适度可使用相同的量表（见表1、2、3）。

### 2.3.3 PC法的意见分数据

PC法是以偏好进行判断的。



BT.2021-04

## 2.4 单刺激连续质量评价（SSCQE）法

即便在很短的数字编码立体视频片段中，图像质量、深度质量和视觉舒适度也会随时间出现很大变化；这种波动可能取决于场景内容和影响上述三个基本角度的伪影的持续时间（如短期或长期）。SSCQE法的目的就是分析这些动态变化的影响。在SSCQE法中，对立体视频序列的图像质量、深度质量和视觉舒适度进行连续评估（即，按它们随时间的变化）。这种方法被广泛认为更能体现实际的家庭观看模式。

### 2.4.1 测试协议的一般形式

应向被试提供下述格式的测试阶段：

- 节目段（*PS*）：一个节目段对应按某一待评质量参数（*QP*）（例如比特率）处理的一种节目类型（例如体育、新闻、戏剧）；每个节目段应持续至少5分钟。
- 测试阶段（*TS*）：一个测试阶段是由一种或多种不同的*PS/QP*组合构成的一个系列，其中没有间隔且按伪随机顺序排列。每个测试阶段至少有一次含有全部*PS*和*QP*，但不必含有全部的*PS/QP*组合；每个测试阶段的长度应介于30至60分钟之间；
- 测试演示（*TP*）：一个测试演示是对某次测试的全面执行。一个测试演示可分成若干测试阶段，以便符合最长持续时间要求，并针对所有*PS/QP*对进行质量评价。如果*PS/QP*对的数量有限，测试演示可重复相同的测试阶段，以便在足够长的时间段内进行测试。

对于服务质量评价，可引入音频。在这种情况下，在进行测试之前，伴音素材的选择应被视为与视频素材选择具有同等的重要性。

最简单的测试格式是使用单一的节目段和单一的质量参数。

### 2.4.2 SSCQE法的分级量表

对于图像质量评价和深度质量，应使用国际电联标准的连续质量量表（见表1和表2）。对于视觉舒适度，应使用表3所示的连续舒适度量表。表5是使用质量量表的测试阶段示例。

### 2.4.3 SSCQE法的意见分数据

收集整理所有测试阶段的数据，计算平均质量评分，作为时间的函数 $q(t)$ 。结果可以每个节目段、视频内容或测试阶段所有观察者的质量评分的平均值表示。

图 5

SSCQE法典型测试阶段输出结果

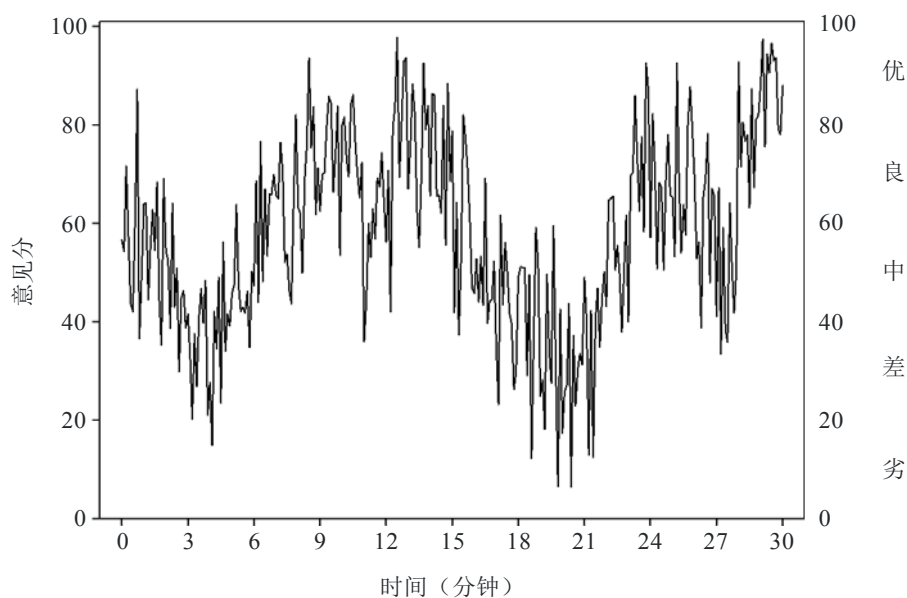


表 1

图像质量的主观评价方法




演示模式	序列持续时间	二进制量表	非连续量表	连续量表
ITU-R BT.500建议书附件1第6.1段所述的单刺激(SS)法	~10 s		5 优 4 良 3 中 2 差 1 劣	 优 良 中 差 劣
ITU-R BT.500建议书附件1第5段所述的双刺激连续质量量表(DSCQS)法	~10 s			 优 良 中 差 劣
ITU-R BT.500建议书附件1第6.2段所述的刺激比较(SC)法	~10 s	A vs. B	-3 甚差 -2 差 -1 稍差 0 相同 1 稍好 2 较好 3 甚好	
ITU-R BT.500建议书附件1第6.3段所述的单刺激连续质量评价(SSCQE)法	~3-5分钟			 优 良 中 差 劣

表 2

深度质量的主观评价方法




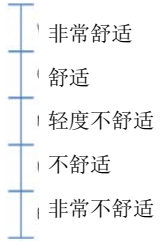
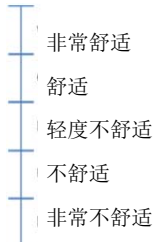
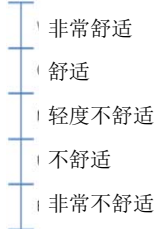
演示模式	序列持续时间	二进制量表	非连续量表	连续量表
ITU-R BT.500建议书附件1第6.1段所述的单刺激(SS)法	~10 s		5 优 4 良 3 中 2 差 1 劣	 优 良 中 差 劣
ITU-R BT.500建议书附件1第5段所述的双刺激连续质量量表(DSCQS)法	~10 s			 优 良 中 差 劣
ITU-R BT.500建议书附件1第6.2段所述的刺激比较(SC)法	~10 s	A vs. B	-3 甚差 -2 差 -1 稍差 0 相同 1 稍好 2 较好 3 甚好	
ITU-R BT.500建议书附件1第6.3段所述的单刺激连续质量评价(SSCQE)法	~3-5分钟			 优 良 中 差 劣

表 3  
视觉舒适度的主观评价方法

演示模式	序列持续时间	二进制量表	非连续量表	连续量表
ITU-R BT.500建议书附件1第6.1段所述的单刺激 (SS) 法	~10 s		5 优 4 良 3 中 2 差 1 劣	
ITU-R BT.500建议书附件1第5段所述的双刺激连续质量量表 (DSCQS) 法	~10 s			
ITU-R BT.500建议书附件1第6.2段所述的刺激比较 (SC) 法	~10 s	A vs. B	-3 甚差 -2 差 -1 稍差 0 相同 1 稍好 2 较好 3 甚好	
ITU-R BT.500建议书附件1第6.3段所述的单刺激连续质量评价 (SSCQE) 法	~3-5分钟			

### 3 一般观看条件

观看条件（包括屏幕亮度、对比度、背景亮度、观看距离等）应与ITU-R BT.2022建议书（6/20号文件）（平板监视器的SDTV和HDTV电视图像主观质量评估的通用观看条件）中所述用于2D的那些观看条件一致。这种一致性方法的理由有两方面。其一，在实际中用户以与2D相同的显示器和观看条件观看3DTV。其二，3DTV视频技术性能的进展通常需要与标准HDTV视频技术的进展联系起来（即对比）衡量。

ITU-R BT.2022建议书（6/20号文件）规定了两种可能的观看距离选择标准。这里将选择设计观看距离（DVD）。对数字系统而言，DVD是指两个邻近像素对应观看者眼睛形成1弧分角度的距离。



如以图像高度的倍数表示，对于1280 x 720（ITU-R BT.1543和ITU-R BT.1847建议书）图像分辨率的系统，DVD为4.8H；对于1920 x 1080家庭（ITU-R BT.709建议书）HDTV图像分辨率系统为3.1H（静态图像）。

为了便于说明，表4描述了有代表性的电视机对角线尺寸的设计观看距离（米）。

表 4  
不同电视机对角线尺寸的设计观看距离

对角线尺寸（英寸）	1920 × 1080图像系统	1280 × 720 图像系统
	设计观看距离（米）	设计观看距离（米）
32	1.24	1.88
42	1.62	2.47
52	2.01	3.06
62	2.39	3.64
72	2.78	4.23
82	3.17	4.82
92	3.55	5.41
102	3.94	5.99

应注意，两个邻近像素对应观看者眼睛成1弧分的角度，那么在设计观看距离，系统可描绘出的最小角度（视网膜）像差（即系统的深度分辨率）等于1弧分（或60弧秒）。研究表明，近97%的人能够区分等于或小于140弧秒的水平像差，至少80%能察觉30弧秒的水平像差。因此，大多数观看者应该能在设计观看距离分辨目前3D视频系统可描绘出的最小像差。

#### 4 测试素材

应按研究中要解决的实验问题选择测试素材。一般情况下，测试序列在内容（体育、戏剧、电影等）及其时空特性上应在被研究业务所传送的节目中具有代表性。

此外，选定的立体测试序列内容通常亦应可以舒适地观看。立体图像的视觉舒适度主要取决于图像包含的像差（视差）和观看条件。因此，应注意确保这种像差不超过下一节中列出的限值，除非研究的目的是专门衡量视觉舒适度。此外，在可能的情况下，应测量和报告测试序列像差分布的统计数据：平均值、标准偏差和范围（最小/最大）。

在选择易于观看的立体3D图像的测试图像时，可将视差（即左右眼图像之间的不一致性）以及视差的分布和变化列为应考虑的项目。易于观看的立体3D图像与视差（左右眼图像之间的不一致性）以及视差的分布和变化之间的关系见随后几个分节的内容。

#### 4.1 视觉舒适度限值

像差/视差过大会导致视觉不舒适，原因可能是这会恶化调节与聚焦功能之间的冲突。因此已提出建议，如要尽可能减小调节与聚焦之间的冲突，那么立体图像的像差应足够小，以便感知的物体深度在“舒适区”范围内。已提出了若干方法来界定这些限值。一种方法是用屏幕视差的程度来规定舒适观看的限值，以屏幕水平尺寸的百分比表示。并建议对于交叉/负像差，这一限值为1%，对于非交叉/正视差，为2%（总计约为3%）。而另一种方法是，舒适区根据眼睛的景深界定。对于电视广播典型的观看条件，研究人员已假定景深介于0.2D（屈光度）与±0.3D（屈光度）之间。对于从3.1H的设计观看距离观看1920×1080（ITU-R BT.709建议书）HDTV图像分辨率的系统，这些值约对应屏幕视差的±2%和±3%。最后，另一种方法从视网膜像差角度规定舒适度限值，并设定对于正负像差，这些限值均为视角的±1°。

值得一提的是，这些不同的方法往往会得到相同的舒适度限值。回想在设计观看距离，两个邻近像素对应观看者眼睛成1弧分的角度。因此，60个像素对应1°的视角。这样我们就可以很容易地从视网膜像差角度规定舒适度限值（对一般观看者）。例如，对于1920×1080（ITU-R BT.709建议书）HDTV图像分辨率的系统，1%（~19.2个像素）约对应20弧分，2%约对应40弧分，3%约对应60弧分（或1°）。

应注意，尽管在设计观看距离，两个邻近像素总是对应1弧分的角度，二者之间的物理分隔距离（如以毫米为单位）随显示器的增大（像素的数目不变，但屏幕的物理尺寸增加）而增加。因此，对较大的显示器，较高的限值（如±3%）会导致对应点之间的物理距离（两个视图的视差，以毫米为单位）超过一般观看者的瞳距（~63-65毫米）。这可能会增加不舒适感。

#### 4.2 左右图像之间的差异

在立体3D系统中，通过将左右图像分别呈现给左右两眼而形成双目3D图像。如两个图像之间产生差异，可导致身心应激，在一些情况下可能无法观看3D图像。例如，在拍摄和播放立体3DTV节目时，左右图像之间可能会出现几何畸变，如大小不一致、垂直移位、旋转误差等。测试图像最好不要出现这些几何畸变。欲了解进一步的信息，见ITU-R BT.2160-2号报告附件4的第3.2.1段。

在选择易于观看的立体3D图像的测试图像时应考虑的关于左右图像之间差异的项目如下：

- 几何差异，包括大小、垂直位移和偏转；
- 亮度差异，包括黑白度；
- 串扰。

#### 4.3 视差的范围、分布和变化

视差分布与立体图像的视觉舒适度相关。

立体图像的视差分布在场景切换帧期间是不连续的。极端视差或视差突变会造成视觉不适感，因此小心控制测试图像的视差甚为重要。欲了解进一步的信息，见ITU-R BT.2160-2号报告附件4第3.2.2段。

一般情况下，鉴于使用立体测试序列的研究可能会引起某种程度的视觉不适感，因此推荐在可能的情况下使用像差不超过舒适度限值的测试素材，虽然偶尔超过这些限值的情况可能是允许的。

## 5 实验装置

实验装置（视频服务器、显示器等）应能够播放全分辨率HD测试序列，如使用HDMI帧封装格式。这可使可开展的研究的范围具有更大的灵活性。

至今尚未规范用于3DTV评估的参考显示器。因此，大多数研究人员预计将使用目前的消费级3DTV显示器。鉴于这类显示器的特性可能因制造商不同亦各不相同，大力提倡研究人员描述研究中使用的显示器的设置信息。

## 6 观察者

### 6.1 样本容量

一般情况下，推荐使用至少30名观看者。但人们认识到，既然3D研究的样本容量考量与2D研究不同，那么实际的数量将取决于研究的具体目标。

### 6.2 视力筛选

应使用目前的临床视力测试方法从视敏度、色盲和立体视觉方面对观察者进行筛选（如对视敏度使用斯耐伦（Snellen）视力表，针对色觉使用Ishihara（石原）假同色图或等效方法，针对立体视觉使用Rando立体图或等效方法）。注意，Randot、Stereo Fly或Frisby测试等立体视测试通常测量从约20弧秒至400弧秒的视网膜像差。提倡研究人员描述参与研究的观察者立体视力方面的相关统计数据。如需对参与者的立体视力进行更详细的分析，研究人员可使用附录1中所示的测试素材。

## 7 观察者须知

应针对研究角度（如深度质量、舒适度等）编写须知。值得一提的是，3D研究的道德准则比2D图像质量评估通常使用的准则更为严格，因为参与者可能会出现视觉不适。一般情况下，3D研究中在向参与者说明研究动机以及暴露于研究中使用的刺激要素可能产生的任何负面影响时要更为谨慎。

## 8 测试阶段持续时间

如观看素材被认为是舒适的，那么测试阶段的持续时间可能与2D研究一样长（即加休息时间~20-40分钟）。如已知素材视差过大，因此存在已知的不舒适的可能性，那么应对持续时间加以限制。

## 9 参考视频素材的使用

研究人员可能希望将参考序列（如有的话）纳入测试序列集。参考序列通常是测试序列未经任何处理的版本（即原始的源序列）。对于立体研究，主要的参考是原始的、未经处理的立体序列。但实验计划可能还包含参考序列的单视场版（即原始源序列的单视图版）；例如，在视觉舒适度研究中，将单视场参考序列的视觉舒适度作为基准可能是有帮助的。单视场参考序列应以3D模式演示（如使用与实际的立体序列相同的3D软件设置对左右两眼演示左视图）。在实验计划中纳入参考序列有两个重要优势。其一，提供了衡量被研究算法或技术实现的透明度（亦称保真度）的机会<sup>2</sup>。其二，纳入参考序列就相当于提供了一个高质量的参照，可以帮助稳定观察者的评分<sup>3</sup>。

## 10 响应的可变性

主观评价实验中观看者提供的评分通常是多变的。观看者之间的差异可能只反映了参考人群的特性，因此可通过增加样本容量解决这个问题。

但是，其中部分可变性可能源自实验过程中个体观看者响应模式的变化。这些变化意味着评价标准的改变，由于任务练习的增加以及熟悉伪影特性等情况，这种变化是有可能发生的。为了尽可能降低这种可变性的负面效应，研究人员应提供适当的培训程序（任务、退化程度等），使用多种随机化方式（即以不同的随机顺序向不同的观看者演示测试序列），并重复试验（通过这种方式亦可衡量响应模式可能的变化）。

## 11 观看者的舍弃标准

针对第2段所述的方法，观看者的舍弃标准见ITU-R BT.500建议书。

## 12 统计分析

3D成像系统研究的统计分析与2D成像系统相同。

---

<sup>2</sup> 透明度（保真度）是描述与毫无退化条件下理想传输系统相关的编解码器或系统性能的概念。显而易见，可通过比较参考序列的得分与经过被研究算法或技术处理过的序列的得分，衡量透明度。

<sup>3</sup> 人们认识到，可能亦可通过低质量的参照提高跨空间（即在不同实验室）和时间（即在同一实验室不同时间）的评分稳定性。但国际电联确实计划立即着手为立体成像技术评价制定/定义标准化的低质量参照。

## 附录 1

## 视力测试的测试素材

## 1 视力测试

表5列出了视力测试的测试用图。这12种测试是根据人类视觉系统的层次结构从低到高选定的。下文描述了八种主要的视力测试（VT），其他四种是临床测试。观察者必须有正常的立体视觉，这意味着他们必须通过精细立体视VT-04和动态立体视VT 07测试。其余六种测试则针对更细微的表征。测试图应放在显示器屏幕高度的三倍处观看。

表 5

视力测试的立体测试素材

编号	项目	测试目的	内容
1	同时视	同时感知两眼分视图像且合成位置正确的能力	呈现给一只眼的图像是一个笼子，呈现给另一只眼的图像是一头狮子
2	双眼融像	将左右眼分视图像合二为一的能力	呈现给一只眼的图像有两个点，呈现给另一只眼的图像有三个点，其中一个点是共同的
3	粗略立体视	将有视差的两眼分视图像合成为一个有粗略深度的图像的能力	呈现给两眼的图像是一只翅膀展开的蜻蜓的立体像对
4	精细立体视	将有视差的两眼分视图像合成为一个有精细深度的图像的能力	提供九个测试用菱形图形，每个菱形中有四个圆圈，其中一个略有视差
5	交叉融像限值	将有交叉像差的两眼分视图像合二为一的能力	演示长条图形的立体像对，交叉视差变化率为10'/s
6	非交叉融像限值	将有非交叉视差的两眼分视图像合二为一的能力	演示长条图形的立体像对，非交叉视差变化率为11'/s
7	动态立体视	在移动随机点立体图像中感知深度的能力	动态随机点立体图
8	双目锐度	双目锐度，包括可能影响良好立体视觉的单眼锐度不平衡	各种方向和大小的E字符
9	水平斜视	患者无法克服的眼位水平偏斜	垂直线和水平线
10	垂直斜视	患者无法克服的眼位垂直偏斜	垂直线和水平线



表 5 (完)

编号	项目	测试目的	内容
11	不等像	两眼视像的形状和大小不一样的情形	左眼用图包括“[o”字符，右眼用图包括“o]”字符，其中字符“o”的位置是共同的
12	旋转隐斜	融合作用遭到阻断时一只眼绕垂直轴的偏斜趋势	左眼用图是钟表的表盘，右眼用图是指向六点钟的钟表指针

注 1 – 这些素材采用1125/60/I 格式（见ITU-R BT.709建议书）。

注 2 – 这些素材可从图像信息与电视工程师学会（ITE）获得，地址3-5-8 Shibakoen, Minato-ku, Tokyo 105-0011, Japan，电话：81-3-3432-4675，电子邮件：[ite@ite.or.jp](mailto:ite@ite.or.jp)。

下面用并排摆放进行交叉自由融像的左右缩略图加以说明。

### 1) VT-01：同时视（狮子测试）

测试同时感知两眼分视图像且合成位置正确的能力。呈现给一只眼的图像是一个笼子，呈现给另一只眼的图像是一头狮子，狮子位置变化速率为12'/s。每个图像的大小固定在10°，以便观察者可以在旁黄斑区捕获图像。视力正常的观察者可以在演示阶段内的某个时间看到狮子在笼子里。

图 6

VT-01测试图



BT.2021-06

### 2) VT-02：双眼融像（worth 4点测试）

测试将左右眼分视图像合二为一的能力。呈现给一只眼的图像有两个点，呈现给另一只眼的图像有三个点，其中一个点是共同的。视力正常的观察者可以看到四个点。

图 7  
VT-02测试图

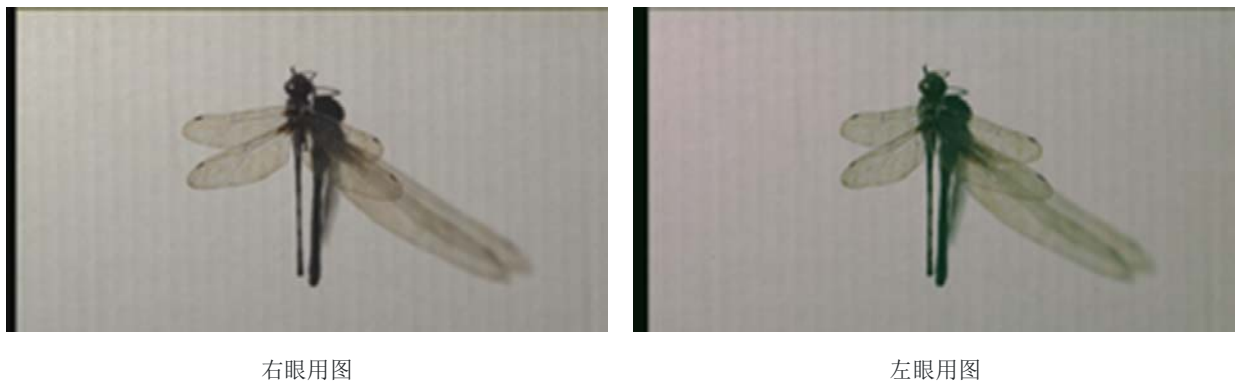


BT.2021-07

### 3) VT-03: 粗略立体视 (蜻蜓测试)

测试将有视差的两眼分视图像合成为一个有粗略深度的图像的能力。呈现给两眼的图像是一只翅膀展开的蜻蜓的立体像对。视力正常的观察者可看到蜻蜓的翅膀在显示屏前面。

图 8  
VT-03测试图



BT.2021-08

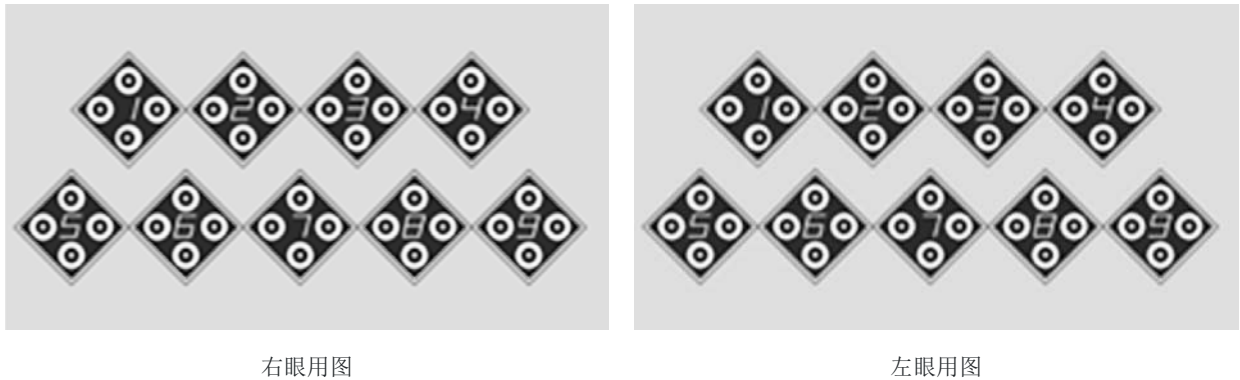
### 4) VT-04: 精细立体视 (圆圈测试)

测试将有视差的两眼分视图像合成为一个有精细深度的图像的能力。提供九个测试用菱形图形，每个菱形中有四个圆圈，其中只有一个略有视差。视力正常的观察者可看到具有视差的圆圈在显示屏的前面。表6列出了测试编号、正确答案和3 H处的立体角度。

表 6  
正确答案和视差

测试编号	正确答案	3 H处的立体角度 (")
1	下	480
2	左	420
3	下	360
4	上	300
5	上	240
6	左	180
7	右	120
8	左	60
9	-	0

图 9  
VT-04测试图



BT.2021-09

### 5) 交叉融像限值（长条图形测试）

测试将有交叉像差的两眼分视图像合二为一的能力。演示长条图形的立体像对，视差变化率为10"/s。可以测量上升和下降系列的融像限值。指示观察者在上升系列中一看到两个图像立刻报告融像间断，在下降系列中一看到的分视图像合为一个单一图像时报告其融像恢复。

图 10  
VT-05测试图



右眼用图

左眼用图

BT.2021-10

6) **VT-06: 非交叉融像限值（长条图形测试）**

测试将有非交叉视差的两眼分视图像合二为一的能力。演示的图像与上述交叉测试中的相同，但左右眼用图对调。

图 11  
VT-06测试图



右眼用图

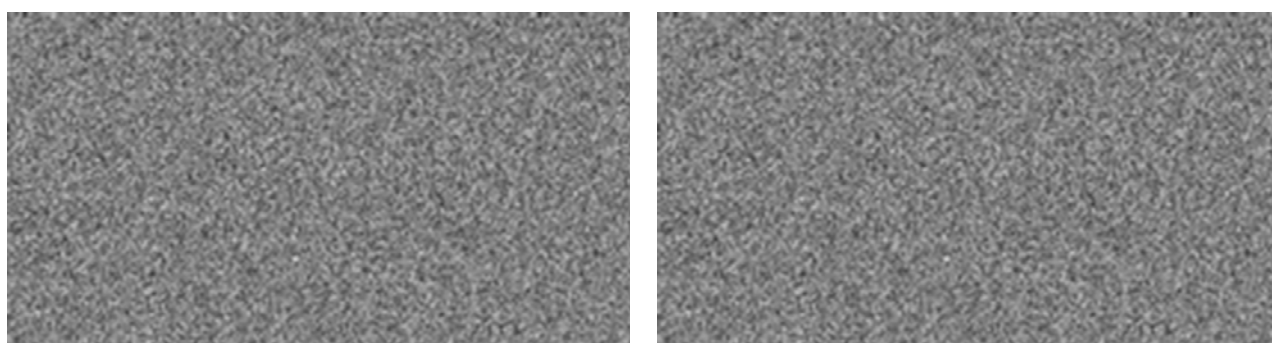
左眼用图

BT.2021-11

7) **VT-07: 动态立体视（动态随机点立体图测试）**

测试在移动随机点立体图像中感知深度的能力。视力正常的观察者可在动态随机点立体图中看到一个矩形的形状和正弦深度运动。

图 12  
VT-07测试图



右眼用图

左眼用图

BT.2021-12

### 8) VT-08: 双目锐度（锐度测试）

测试双目锐度及双眼融像，包括可能影响良好立体视觉的单眼锐度不平衡。图像由四列五行各种方向和大小的E字符组成。中间两列，双眼皆看看到，左边两列仅左眼可以看到，右边两列仅右眼可以看到。视力正常的观察者可以正确说出E字符的方向。字符的大小对应3 H处的锐度：约1.0、0.5、0.33、0.25和0.125。

图 13  
VT-08测试图



右眼用图

左眼用图

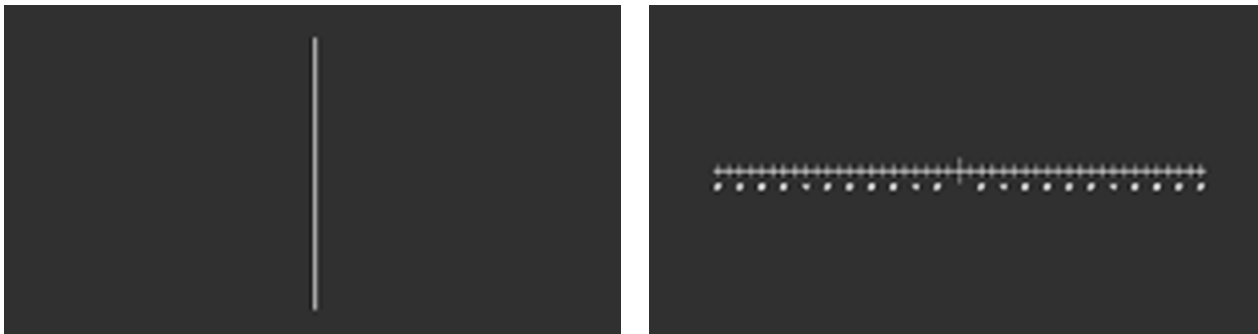
BT.2021-13

### 9 & 10) VT-09: 水平斜视（水平马氏杆（maddox）测试）和VT-10: 垂直斜视（垂直马氏杆测试）

这些图用于衡量眼位的水平和垂直偏斜。双眼视轴假定了一个相对于彼此、与生理条件要求不同的位置。演示的图像包括一条垂直线和一条水平线。视力正常的观察者可以看到两线的交叉点约在二者的中心处。标记旁数字的单位是棱镜度，其中PD（瞳距）=65毫米，观看距离为3.02 H。



图 14  
VT-09测试图



BT.2021-14

图 15  
VT-10测试图



BT.2021-15

### 11) VT-II: 不等像（“[]”字符测试）

两眼视像的形状和大小不一样的情形。左眼用图包括“[o”字符，右眼用图包括“o]”字符，其中字符“o”的位置是共同的。视力正常的观察者可以看到相同大小相同高度的“[”和“]”。

图 16  
VT-11测试图



BT.2021-16

## 12) VT-12: 旋转隐斜 (钟表测试)

只有在一眼被遮盖或融合作用遭到阻断时才会表现出来的绕垂直轴的眼位偏斜。左眼用图是钟表的表盘，右眼用图是指向六点钟的钟表指针。视力正常的观察者可以看到钟表恰好显示六点钟。

图 17  
VT-12测试图



BT.2021-17