



中华人民共和国国家标准

GB/T 22123—2008

数字电视接收设备图像和声音 主观评价方法

Methods for picture and audio subjective assessment
of digital television receiving equipment

2008-06-30 发布

2009-01-01 实施



中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

中 华 人 民 共 和 国
国 家 标 准
数字电视接收设备图像和声音
主观评价方法
GB/T 22123—2008

*

中国标准出版社出版发行
北京复兴门外三里河北街16号
邮政编码:100045

网址 www.spc.net.cn

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 1.25 字数 31 千字
2008年9月第一版 2008年9月第一次印刷

*

书号: 155066 · 1-33730 定价 18.00 元

如有印装差错 由本社发行中心调换

版权专有 侵权必究

举报电话:(010)68533533



GB/T 22123-2008

前 言

本标准的附录 A、附录 B、附录 C 为规范性附录。

本标准由中华人民共和国信息产业部提出。

本标准由全国音频、视频及多媒体系统与设备标准化技术委员会归口。

本标准主要起草单位：国家广播电视产品质量监督检验中心、中国电子科技集团公司第三研究所、厦门华侨电子有限公司、信息产业部电子第四研究所、四川长虹电器股份有限公司、青岛海信电器股份有限公司、深圳康佳集团股份有限公司、TCL 集团股份有限公司。

本标准参加起草单位：北京牡丹电子集团有限责任公司、苏州飞利浦消费电子有限公司、熊猫电子股份有限公司、重庆现代视声公司、上海宽带数码技术有限公司、北京中视联数字系统有限公司、江西科创数显技术中心科技公司、上海凤凰光电有限公司、上海广电金星电子股份有限公司、清华大学、江苏省电子产品监督检验所、深圳创维-RGB 电子有限公司、上海广电股份平面显示器分公司、北京青鸟华光科技有限公司、上海永新彩色显象管股份有限公司、深圳电子产品质量检测中心、上海市电子仪表标准计量测试所、广东省电子产品检验所、深圳市数视通信息技术有限公司、广东东菱电器集团有限公司、广东步步高电子工业有限公司、深圳康佳集团、四川中视联数字设备有限公司、北京市数字高清晰度电视产业工程协调小组办公室、江苏新科电子集团有限公司、联想(北京)有限公司、北京神州数码有限公司、泰克电子(中国)有限公司、信息产业部电子第五研究所、电子 203 计量站。

本标准主要起草人：李剑、王海燕、阮卫泓、吴昕、吴蔚华、陈科、李庆新、王伟、马德彪、宋丹玫、赵燕泥、许福平、滕荣庆、周金忠、汪莉。

引 言

本标准参照国际电信联盟 ITU-R BT. 500-10、ITU-R BT. 710-2、ITU-R BT. 1129-2 以及 ITU-R BT. 1201-1 建议书制定。

由于数字电视信号在传输过程中会受到不同程度的损伤,本标准选用国际推荐的双刺激损伤标度法和刺激比较法作为数字电视接收设备的图像质量主观评价方法。考虑到电视广播中接收端的信号特征,在构成主观评价系统的方案上与 ITU-R BT. 710-2、ITU-R BT. 1129-2 以及 ITU-R BT. 1201-1 建议书的规定有所不同。

目 次

前言 I

引言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 主观评价的一般要求 2

4.1 环境条件 2

4.2 信号源 4

4.3 测试素材的选择 4

4.4 评价员 5

4.5 评价试验阶段 5

4.6 结果的表达 7

5 评价试验方法与数据处理 7

5.1 评价方法 7

5.2 双刺激连续质量标度法 7

5.3 刺激比较法 9

5.4 数据处理..... 10

附录 A（规范性附录） 评价结果的筛选 11

附录 B（规范性附录） 双刺激连续损伤标度法评分登记卡 12

附录 C（规范性附录） 刺激比较法评分登记卡 13

参考文献 14

数字电视接收设备图像和声音 主观评价方法

1 范围

本标准规定了数字电视接收设备的主观评价条件和评价方法。

本标准适用于各种接收和显示方式(如卫星、地面和有线等)的接收器、接收机和显示器(如阴极射线管、液晶、等离子及各种投影等显示方式)等数字电视设备(包括标准清晰度和高清晰度)的主观评价,其他类似用途的设备可参照执行。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 14857 演播室数字电视编码参数规范

SJ/T 11324 数字电视接收设备术语

SJ/T 11348—2006 数字电视平板显示器测量方法

GY 155 高清晰度电视节目制作及交换用视频参数值

3 术语和定义

SJ/T 11324 确立的以及下列术语和定义适用于本标准。

3.1

主观评价 subjective assessment

通过评价人员对被测系统质量的主观反应,来确定电视系统性能的一种方法。

3.2

测试素材 test fodder

用于主观评价电视图像质量和声音质量,并且对图像内容和声音内容有特定要求的图像或图像序列。

3.3

评级标度 grading scale

主观评价中用于评定图像质量和声音质量的评定尺度。

3.4

图像干扰 picture interference

亮度和(或)色度的不规则性或损伤图像稳定性的干扰,或者是它们的任意组合。

注1:通常的亮度和(或)色度的不规则性如:

- a) 从顶到底或从左到右本底亮度的变化;
- b) 在屏幕左边或右边的一层轻微模糊物;
- c) 屏幕左侧的一些垂直窄带;
- d) 在整个屏幕上的垂直宽带;

- e) 背景中有黑色水平带(图像波纹或哼声干扰);
- f) 同时传输的音乐或语言音量的节奏有关的水平条(伴音对图像的影响,又称图像微音效应);
- g) 由信号传输误码引起的“块效应”;
- h) 图像边沿的“飞蚊效应”。

注 2: 破坏图像稳定度的最常见的干扰如:

- a) 整个图像或其局部的垂直与水平抖动;
- b) 图像本应静止的部分出现波状运动;
- c) 图像不规则跳动、静止、黑屏/蓝屏等;
- d) 水平扫描线撕裂;
- e) 由行频变换带来的图像干扰和损伤;
- f) 由帧频变换带来的图像干扰和损伤;
- g) 由块估计产生的振铃效应。

3.5

总图像质量 general picture quality

评价人员对评价过的图像质量的综合印象。

3.6

对伴音的干扰 associated audio interference

在数字电视接收设备内产生的不需要的声音信号,或其部件的机械震动所造成的干扰。

注: 常见的对伴音的干扰如:

- a) 由电源电压滤波不良引起的嗡声干扰;
- b) 垂直偏转频率引起的稳定蜂音;
- c) 由音频信号传输误码引起的声音中断;
- d) 由内部静电放电或误码引起的咔咔声;
- e) 垂直偏转部件或电源变压器的机械震动引起的噪声;
- f) 由行输出高压变压器或偏转电路的其他零件的震动引起的行频啸叫;
- g) 其他零件引起的噪声。

3.7

总伴音质量 general associated audio quality

在注意到刚刚试验过的各种伴音性能的质量之后,评价人员对伴音质量的总的印象,特别是对于柔和的音乐,需重点评价的是声音的清晰度。

3.8

总重现质量 general present quality

在评价人员注意到刚刚试验过的各种图像和伴音性能的质量之后,对接收设备的总的质量印象。

4 主观评价的一般要求

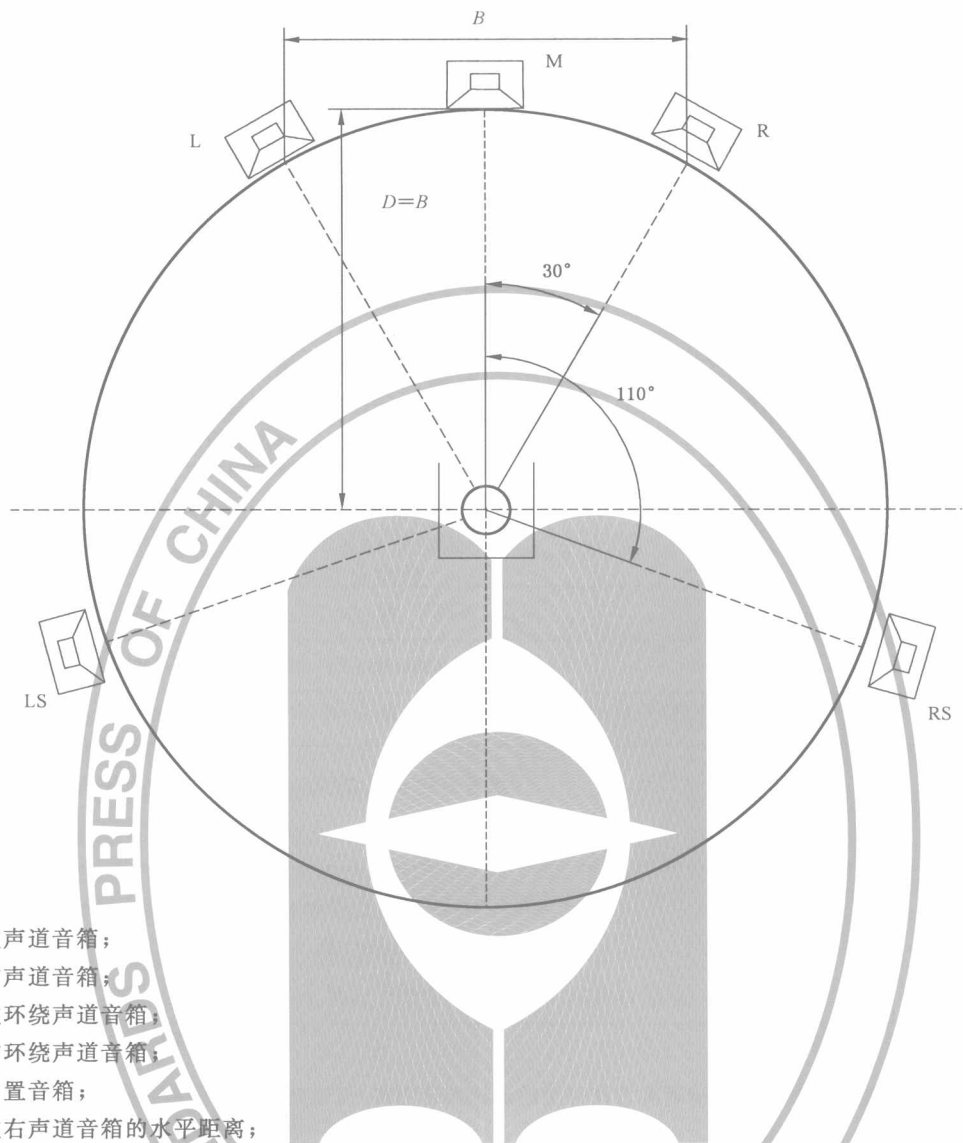
4.1 环境条件

主观评价实验室环境应类似或优于一般家庭住房环境,观看条件应满足表 1 规定的要求,听音评价时评价人员的位置如图 1 所示(见参考文献[5])。

表 1 数字电视接收设备主观评价观看条件

项 目	SDTV 参数值	HDTV 参数值
观看距离 ^a	4~6 倍有效图像高度	2~3 倍有效图像高度
有用平均亮度 ^b	按 4.5.3.4 进行调整,达到各通用规范规定的亮度值	
屏幕全黑屏幕亮度与峰值亮度之比 ^{c,d}	≤0.02	≤0.02
背景亮度与图像峰值亮度之比 ^d	约 0.15	约 0.15
暗室中,黑电平亮度与峰值亮度之比 ^e	约 0.01	约 0.01
室内环境光照明	30 lx ~75 lx	30 lx ~75 lx
背景光光源色温	D ₆₅	D ₆₅
背景光光源亮度	小于被测设备 平均亮度 10%	小于被测设备 平均亮度 10%
背景光部分对评价员的张角 ^{f,g}	高:≥43° 宽:≥57°	高:≥53° 宽:≥83°
背景噪声声级(A 计权)	≤35 dB	≤35 dB
混响时间	0.4 s ~0.5 s	0.4 s ~0.5 s
评价员的座位安排	水平方向在中垂线 ±30°以内	水平方向在中垂线 ±30°以内
电源电压	220(1+10%)V,频率振动 ≤±2%,谐波分量 ≤±5%	220(1+10%)V,频率振动 ≤±2%,谐波分量 ≤±5%
^a 6 倍图像高度是用于数字标准清晰度系统评价的设计观看距离,但也允许在 4 倍图像高度处评价,评价结果可以分别给出。 ^b 当有用平均亮度达不到规定要求时,应注明。 ^c 受室内环境照明和显示对比度范围的影响。 ^d 应设置室内环境照明。 ^e 黑电平对应于 0%的视频信号幅度。 ^f 若达不到规定值时,应注明,且背景光面积最少不得小于被测设备面积的 130%。 ^g 若 HDTV 达不到规定值时,应满足高≥28°,宽≥48°。		

听音评价时(见参考文献[4]),若数字电视接收设备为 5.1 声道,各音箱的摆放位置如图 1 所示,评价人员与各声道音箱的距离 D 应该和左右音箱的水平距离 B 相等。



- L——左声道音箱；
R——右声道音箱；
LS——左环绕声道音箱；
RS——右环绕声道音箱；
M——中置音箱；
B——左右声道音箱的水平距离；
D——各声道音箱与评价人员的距离。

图 1 听音评价位置图

4.2 信号源

信号源符合 GB/T 14857 和 GY 155 标准规定。测试信号经压缩编码生成的测试码流送至测试解码器，其输出的基准图像由图像监视器显示，声音由外置音箱输出。测试码流经数字电视发射机后，由被测数字电视接收设备接收或显示被测图像，声音由被测数字电视接收设备的音箱或外置音箱输出。

测试信号编码一般采取高码率编码，同时在个别测试信号上兼顾典型码率编码和低码率编码。

4.3 测试素材的选择

选择主观评价测试素材(见参考文献[3])，并应满足以下要求：

- a) 测试图像的色彩、形式、细节、亮度和对比度等不应使评价员分心；
- b) 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，还应包含不同的运动形式如平移、推拉或旋转，每个序列大约持续 10 s~15 s；
- c) 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、彩色再现和运动再现等具有敏感特性的评价因素；
- d) 测试图像应包含肤色、天蓝色、草绿色、红旗色等常见色和能代表电视节目的典型内容；

- 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。例如室内、室外远景、中景、近景、人物、景色、多细节、中等细节、少细节和背景复杂、简单、亮、暗等;
- f) 测试图像应包含多种色调和彩色饱和层次;
- g) 在总的图像序列中最少有一个序列展示刚刚可察觉的源噪声,最少有一个序列是由计算机生成的(见参考文献[1]、[2]);
- h) 临界素材的选择应有固定字长场内编解码素材、固定字长帧间编解码素材、可变字长场内编解码素材、可变字长帧间编解码素材;
- i) 用于听音评价的节目段包括:语声、声乐、器乐和效果声四大类节目素材。

4.4 评价员

- a) 评价员应具有收看、收听经验,并能判断电视图像和声音质量的优劣;
- b) 评价员应具有代表性,即应包括不同性别、年龄、文化程度的人员,他们应具有正常的视力(含矫正视力)和彩色视觉,能较快地接受并掌握评价方法,具有独立的分析判断能力;
- c) 根据不同的评价目的,评价小组可包括专业人员或非专业人员。专业人员是指那些专门从事电视技术和那些对应用于电视的彩色具有专门知识的人。所有其他人员均认为是非专业人员。为消费者进行的试验,评价小组最好由非专业人员组成,可以包括少数专业人员;
- d) 为得到较高置信度的评价结果,评价小组的人数应不少于 15 人,为满足评价环境要求,应考虑分批进行评价试验。

4.5 评价试验阶段

4.5.1 评价示范训练

每次评价试验开始,评价试验负责人应结合评价试验示范演示,向评价人员详细地介绍评价方法、可能存在的图像和伴音质量损伤类型、图像显示顺序和时间、评价制、评分等级和受试设备类型等。其语气应无偏向,不应带有影响评定的暗示。示范说明应事先录音,每次试验开始时播放,以免每次口头说明有出入或有暗示。

示范演示不应使用正式的测试图像,但应与正式的测试图像具有可比性。

4.5.2 评价试验的一般流程

一个包括示范说明在内的评价试验全过程通常不超过 2 h。在评价试验正式开始之前,需要引入 3 个~5 个测试周期来稳定评价员的判断,其评价数据不纳入评价结果的统计中。不应该要求评价小组不停地连续进行 20 min 以上的评价工作,评价小组中间休息的时间至少等于他们进行评价工作的时间。当评价小组第一次工作时,建议有 2 h 的预演时间。

测试图像的显示顺序采用伪随机方式。为了检测相关性,有些测试可以重复进行,但要避免相同测试图像在连续的测试周期出现。评价试验阶段的显示流程如图 2 所示。

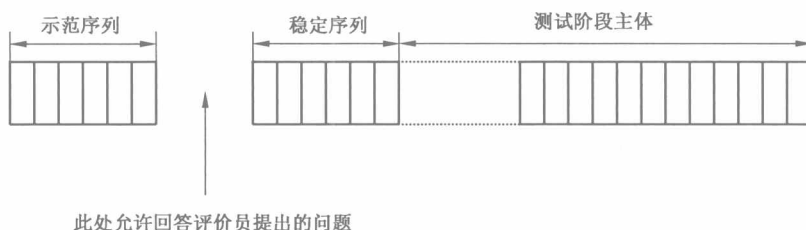


图 2 评价试验阶段的显示流程图

正式开始评价时,评价员应按有关评分等级的说明对正式被测图像和伴音在登记卡上独立地进行记分,试验过程中评价员应集中注意力,防止环境干扰对评价员的视觉和听觉产生影响。

4.5.3 评价试验

4.5.3.1 视频信号

数字电视接收设备进行图像调整时,使用下列视频信号。

4.5.3.1.1 极限八灰度和复合测试图

4.5.3.2 信号输入的要求

数字电视测试发射机的输出信号送至数字电视接收设备的射频输入端,其射频电平和载噪比 C/N 按各传输方式下相应产品国家标准和行业标准规定的标准测试条件进行设置。

4.5.3.3 接收设备的调谐

被测接收设备按相应标准进行设置,正确调谐。

4.5.3.4 接收设备的图像调整

- a) 使用极限八灰度(见 SJ/T 11348—2006 的 4.2.1.6)图像调整接收设备的亮度、对比度调节键,使屏幕上的八级灰度在确保屏幕对比度最大时刚刚可辨;然后调整亮度调节键,使屏幕亮度达到或接近表 1 规定的观看条件。若被评价的某些显示设备要求的数值与规定不同,则所选值及选用原因应在试验结果中加以说明;
- b) 使用复合测试图(见 SJ/T 11348—2006 的 4.2.1.2)调整显示设备的饱和度及色调控制,使屏幕上得到最佳的彩色重现,且让评价人员主观感受舒服自然;
- c) 如果接收设备有清晰度、锐度等控制调整时,按复合测试图调整,使多数评价人员感到图像最佳;
- d) 使用复合测试图调整接收设备,使其重显率大于或等于 95%,若达不到规定要求,应在测试结果中说明;
- e) 具有几何失真特性的接收设备应使用格子信号调整接收设备的几何失真使之最小;
- f) 与伴音效果有关的音量与音调控制应参照接收设备相应标准规定进行设置;
- g) 具有图像改善、声音增强等功能的其他有关控制,应按相应接收设备的需求进行合理设置和调整。

4.5.3.5 接收设备的可见度

为了避免任何心理影响,评价人员仅可观看接收设备的图像,其余部分(如商标、型号等可知接收设备的信息)应加以遮蔽,但不应影响声音效果。

4.5.3.6 图像监视器

测试使用的图像监视器应与被测设备显示方式相同。测试使用的图像监视器应与被测设备显示尺寸、幅型比一致。

图像监视器用作显示基准图像,与被测设备相比较,应具有更优秀的显示特性和稳定的性能,以保证主观评价试验的正确性和可比性。

4.5.3.7 评价

数字电视接收设备图像和伴音质量主观评价的具体内容如下所列,针对不同的测试类型,可选择与测试相关的所有项目进行评价。

- a) 马赛克效应;
- b) 误码造成的亮带或暗带;
- c) 彩色溢出;
- d) 飞蚊效应;
- e) 冻结;
- f) 伪像(包括摩尔效应和格式变换产生的差拍);
- g) 残像;
- h) 拖尾;
- i) 闪烁;
- j) 色纯;
- k) 亮色时延差;

- l) 动态白平衡;
- m) 清晰度;
- n) 图像层次感;
- o) 运动图像的重现性;
- p) 几何失真和图像重显率;
- q) 唇音同步;
- r) 声像定位;
- s) 字幕效果;
- t) 临场感;
- u) 总图像质量;
- v) 总伴音质量;
- w) 总重现质量。

4.6 结果的表达

4.6.1 双刺激连续质量标度法结果的表达

对每个评价序列,应计算出评分的统计平均值、标准偏差和 95%置信度区间。结果还应包括以下内容:

- a) 受试接收设备型号、规格等信息;
- b) 主观评价所采用的方法;
- c) 评价系统使用的设备;
- d) 测试素材的情况及图像源的类型;
- e) 评价员人数及类型;
- f) 测试总平均分;
- g) 如果淘汰了某些评价员,还要说明每个测试序列原始的和调整后的平均分和 95%置信度区间。

4.6.2 刺激比较法结果的表达

- a) 受试接收设备型号、规格等信息;
- b) 主观评价所采用的方法;
- c) 评价系统使用的设备;
- d) 测试素材的情况及图像源的类型;
- e) 评价员人数及类型;
- f) 测试总平均分;
- g) 总图像质量、声音质量的描述。

5 评价试验方法与数据处理

5.1 评价方法

本标准采用双刺激连续质量标度法和刺激比较法两种方法作为主观评价方法,在对数字电视接收器的主观评价中,使用双刺激连续质量标度法;对数字电视接收机的主观评价中,使用刺激比较法。

数字电视接收器主观评价采用双刺激连续质量标度法,数字电视接收机和显示器主观评价采用刺激比较法,其他接收设备参照执行。

5.2 双刺激连续质量标度法

5.2.1 系统框图

测试系统的一般安排如图 3 所示。

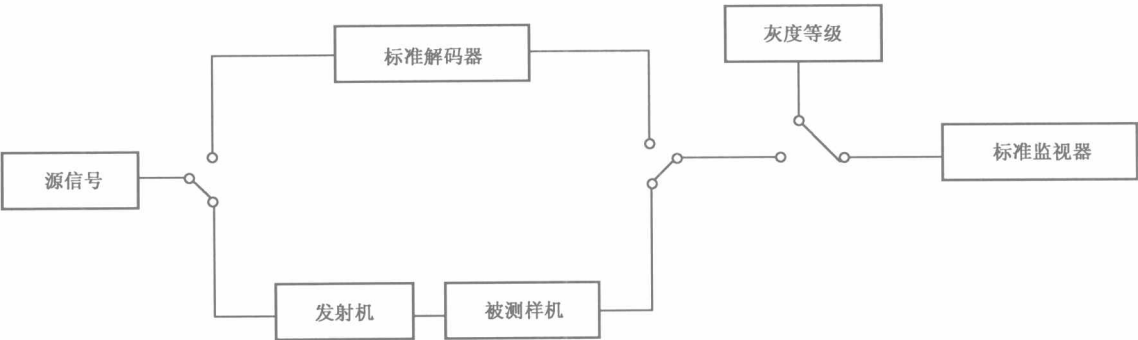


图 3 双刺激连续质量标度法的一般安排

5.2.2 评分标度

双刺激连续质量标度法只要求评价人员在垂直标度上插入一个标记,对每个显示的总图像质量进行评价。为了适应每个图像的双显示,成对地印刷垂直标度。为了避免量化误差,标度提供连续的等级系统,但是它们被划分为五个相等的长度,与正常的 ITU-R 五级质量标度相对应,见图 4 和表 2。

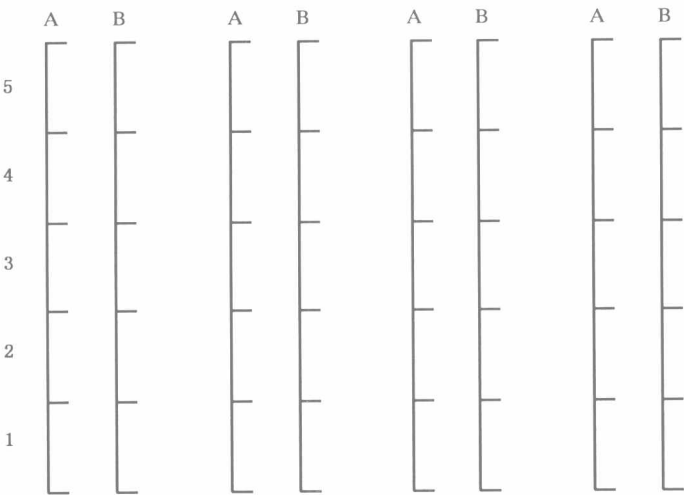


图 4 双刺激连续质量标度的质量等级表格

表 2 质量等级

5	很好
4	好
3	比较满意
2	不太满意
1	不满意

5.2.3 试验图像显示和评分顺序

测试图像显示和评分顺序如图 5 所示。

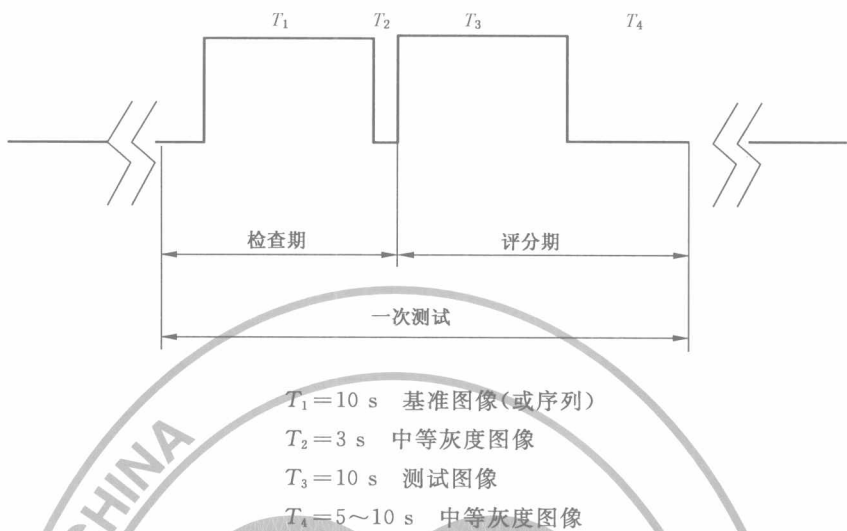


图 5 双刺激连续质量标度法图像显示和评分顺序

5.2.4 结果的解释

从双刺激连续质量标度法得到的结果,不应当以绝对得分来对待,而是应当作为基准条件与测试条件之间的得分差别来对待。因此,把得分与单个质量描述术语联系起来,甚至与那些来自双刺激连续质量标度法协议本身的术语(例如,优、良、中)联系起来,是错误的。

5.3 刺激比较法

5.3.1 一般安排

测试系统的一般安排如图 6 所示。

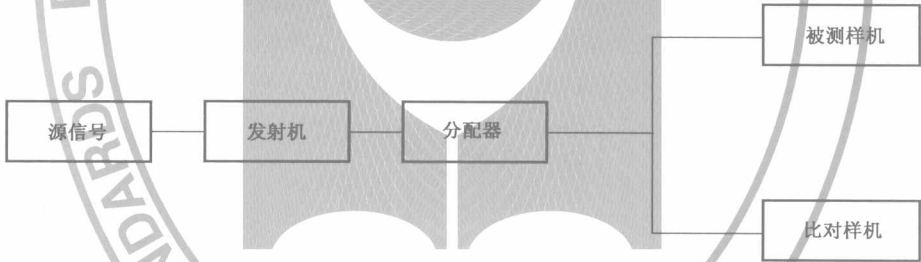


图 6 刺激比较法的一般安排

5.3.2 评分标度

刺激比较法使用形容词范畴判断,这些范畴一般是以语意术语定义的。这些范畴可以报告可察觉的差别的存在(例如,相同、不同),可察觉的差别的存在和方向(例如,较少、相同、较多)或者程度和方向的判断。其比较标度相应的术语在表 3 中给出,评价员根据测试情况,按规定的评分等级给予评定。

表 3 比较标度

-3	坏得多
-2	坏
-1	稍坏
0	相同
1	稍好
2	更好
3	好得多

5.4 数据处理

5.4.1 双刺激连续质量标度法数据处理

将结果(基准图像和被测试的实际图像之间的评级差别)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。
计算出每个被测图像的平均分、标准偏差和 95%置信度区间。
平均分计算公式如下：

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

标准偏差计算公式如下：

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

95%置信度区间如下：

$$(\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} + \delta_{jkr}) \dots\dots\dots (3)$$

式中：

- $\delta_{jkr} = \frac{1.96S_{jkr}}{\sqrt{N}}$;
- u_{ijkr} ——评价员 i 的评分;
- j ——包括基准在内的测试条件的数目;
- k ——测试图像或序列的数目;
- r ——重复次数;
- N ——评价人员的数目;
- S ——标准偏差。

5.4.2 刺激比较法数据处理

计算出每个被测图像的平均分,见公式(4)。

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \dots\dots\dots (4)$$

式中：

- $\bar{x}$ ——每个被测图像的平均分;
- x_i ——评价人员 i 对每个被测图像的评分;
- N ——评价人员的数目。

针对每个被测图像的平均分 $\bar{x}$,按照表 3 的比较标度,与基准图像相比较,给出相应的评价结果。

附 录 A
(规范性附录)
评价结果的筛选

首先,必须利用 β_2 (通过计算函数的峰度系数,即四阶矩对二阶矩的比) 确定每个测试显示的得分是否正态分布。对于每个测试显示,计算平均分 $\bar{u}$ 、标准偏差 S 和峰度系数 β_2 , β_2 由下式给出:

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

式中:

m ——中间过程量;

x ——取值为 2 或 4;

N ——评价人员的数目;

j ——包括基准在内的测试条件的数目;

k ——测试图像或序列的数目;

r ——重复次数。

如果 $2 \leq \beta_{2jkr} \leq 4$, 则这个分布可以取做正态。

对于每个显示,如果测试显示的得分符合正态分布,则

$$P_{jkr} = \bar{u}_{jkr} + 2S_{jkr} \quad Q_{jkr} = \bar{u}_{jkr} - 2S_{jkr}$$

如果测试显示的得分不符合正态分布,则

$$P_{jkr} = \bar{u}_{jkr} + (\sqrt{20})S_{jkr} \quad Q_{jkr} = \bar{u}_{jkr} - (\sqrt{20})S_{jkr}$$

将每个评价人员的得分 u_{ijkr} 与 P_{jkr} 和 Q_{jkr} 进行比较,每当发现评价人员的得分高于 P_{jkr} , 即 $u_{ijkr} \geq P_{jkr}$, 则 $P_i = P_i + 1$; 每当发现评价人员的得分低于 Q_{jkr} , 即 $u_{ijkr} \leq Q_{jkr}$, 则 $Q_i = Q_i + 1$ 。

最后,计算两个比值,如果 $\frac{(P_i + Q_i)}{j \cdot k \cdot r} > 0.05$, 且 $\left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3$, 则剔除评价人员 i 。

式中:

S ——标准偏差;

u_{ijkr} ——评价人员 i 的得分。

附录 B

(规范性附录)

双刺激连续损伤标度法评分登记卡

评价员座位号

日期

年 月 日

姓 名	性 别	年 龄	文化程度	职 业	备 注