GY

中华人民共和国广播电影电视行业标准

GY/T 298—2016

音频系统小损伤主观评价方法

Methods for the subjective assessment of small impairments in audio systems

（ITU-R BS.1116-3，MOD）

2016－03－18 发布　　2016－03－18 实施

国家新闻出版广电总局　发布

目　　次

前　言

本标准按照GB/T 1.1—2009给出的规则起草。

本标准使用重新起草法修改采用ITU-R BS.1116-3《音频系统小损伤主观评价方法》。本标准与ITU-R BS.1116-3相比变化如下：

——第3章的3.1对应ITU-R BS.1116-3的附录4；

——第8章中允许的节目素材最大电平由ITU-R BS.1116-3第6章中规定的“高于校准电平9dB”修改为符合GY/T 282—2014中规定的“最大真峰值音频电平应不超过-2dB TP”。

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

本标准由全国广播电影电视标准化技术委员会（SAC/TC 239）归口。

本标准起草单位：国家新闻出版广电总局广播电视规划院。

本标准主要起草人：张建东、覃毅力、孙岩、王倩男。

引　言

考虑到：

a) ITU-R BT.500、ITU-R BS.1284、ITU-R BT.710 和 ITU-R BT.811 建议书已经建立了一些对视音频系统质量进行主观评价的方法；

b) 有用信号从源端传输至听众的过程中会产生损伤，一类主观听音测试是对损伤带给听众的“不悦”程度进行评价；

c) 传统的客观测量方法不足以评价先进音频编码系统的声音质量，因此开发了感知质量客观评价方法；

d) 使用标准化的方法有利于测试数据的兼容和交换，以及对测试数据的正确评估；

e) 一些利用心理声学特性的先进数字音频系统的新近出现，尤其是产生小损伤的数字音频系统的出现，需要主观评价方法的改进；

f) ITU-R BS.775 规定的多至 3/2 声道的多声道立体声系统和 ITU-R BS.2051 描述的先进声音系统（无论是否伴随有图像）的出现，需要包括测试条件在内的新的主观评价方法。

建议：

使用本标准规定的测试、评价和报告过程对包括多声道在内的声音系统(无论是否伴随有图像）的小损伤进行主观评价。

进一步建议：

适用于先进声音系统的听音室和重放设备的特性有待于进一步研究，待研究完成时，应根据研究结果更新本标准。

音频系统小损伤主观评价方法

1 范围

本标准规定了音频系统小损伤的主观评价方法。

本标准适用于在电视节目或广播节目的收录、分配、传送和监测等环节，对小损伤节目（与源节目对比，源节目须可获得）或产生小损伤的系统的声音质量进行主观评价，也适用于产生小损伤的编解码器等设备的研究和开发。

2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 6278—2012 声系统设备 概述 模拟节目信号

GY/T 192—2003 数字音频设备的满度电平

GY/T 282—2014 数字电视节目平均响度和真峰值音频电平技术要求（ITU-R BS.1864：2010，MOD）

ITU-R BS.645 用于国际声音节目链路的测试信号和电平计量（Test signals and metering to be used on international sound programme connections）

ITU-R BS.708 演播室监听耳机电声特性测量（Determination of the electro-acoustical properties of studio monitor headphones）

ITU-R BS.775 伴随和不伴随图像的多声道立体声声音系统（Multichannel stereophonic sound system with and without accompanying picture）

ITU-R BS.1284 声音质量主观评价通用方法（General methods for the subjective assessment of sound quality）

ITU-R BS.2051 用于节目制作的先进声音系统(Advanced sound system for programme production）

3 术语、定义和缩略语

3.1 术语和定义

下列术语和定义适用于本标准。

3.1.1

片段 excerpt

适于评价给定被测系统声音质量的个性特征或参数的一段音乐、语音或其他声音信号。

测试片段通常为CD、R-DAT或其他格式的一段声音信号。

3.1.2

属性 attribute

根据给定的口头或书面定义，听音测试活动中可感知的特征。

3.1.3

小损伤 small impairments
必须通过严格控制的听音测试条件和适当的统计分析才能觉察到的相比于源素材声音的微小区别。

3.1.4

条目 item
由被测系统处理过的一段片段。

3.1.5

被测对象 object
被测系统，通常以经过该系统处理后的一些测试片段来代表。

3.1.6

参考 reference
未经被测对象处理过的测试片段，用作损伤测试对比的基准。

3.1.7

隐藏参考 hidden reference
未向评价员标识的参考。

3.1.8

刺激 stimulus
被测对象条目、隐藏参考、参考与一个片段的部分或全部的组合。

3.1.9

评价员 subject
在听音测试中评价刺激的测试人员。

3.1.10

评价小组 listening pannel
在一个听音测试中，给出听音测试数据的评价员的整体。

3.1.11

地点 location
听音测试的执行位置，指听音室的地理位置或评价员在听音室内的位置，是测试要素之一。

3.1.12

盲测 blind test
一种测试方法，在该种测试中，刺激是向评价员提供的唯一信息源。

3.1.13

双盲测试　double blind test

盲测的一种，在该种盲测中，听音测试的组织者和听音测试之间没有不受控制的交互可能。

3.1.14

等级评分　grade

根据给定的标度，一个属性量级的数字表示。

3.1.15

一场测试　session

需要由一位评价员或一个评价小组在一个持续的时间段内评估的整组试验（试验定义见下条）。

3.1.16

试验　trial

一场测试的子集，该子集以一组刺激的重放为开始，以完成对它们的等级评分为结束。

一些定义之间的关系说明见图 1。

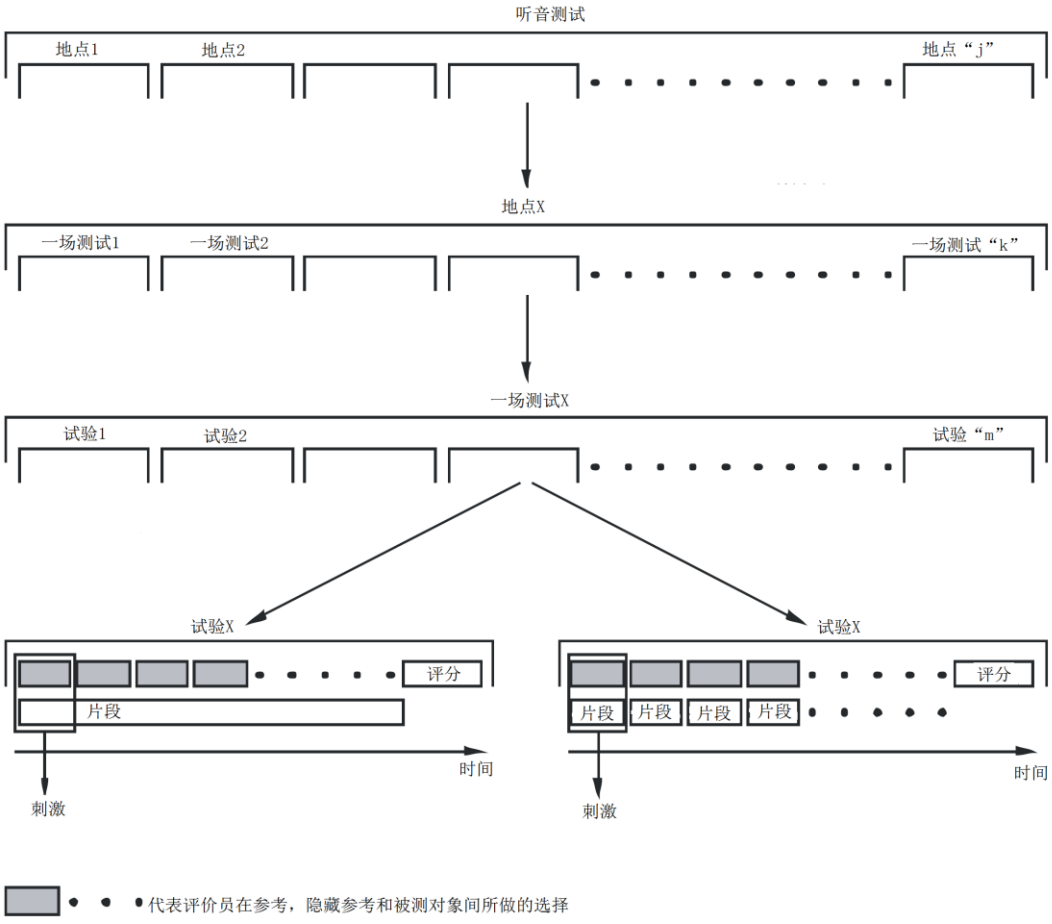

图1　一些定义间的关系

3.2 缩略语

下列缩略语适用于本标准。

ANOVA 方差分析（Analysis of Variance）

SQAM 声音质量评价素材（Sound Quality Assessment Material）

4 测试设计

在科学领域存在很多采集可靠信息的策略。音频系统小损伤的主观评价应采用最严谨的测试方法，首先要严格把控测试条件，其次要把握好评价员的量化数据。

主观测试需要仔细地设计和规划，以避免受到不可控因素的影响而产生歧义。例如，在听音测试中，如果音频条目的实际顺序对所有评价员都相同，则无法确定评价员所给出的判断是出于播放顺序还是出于不同的损伤等级。因此，必须以揭示独立因素且只包括这些因素影响效果的方式安排测试条件。

在整个听音测试中：如果潜在损伤和其他特性预计呈均匀分布，则可以采用真正随机化的方式安排测试条件；如果预计呈非均匀分布，则必须考虑测试条件的安排方式。例如，如果待测素材的难易程度不同，则刺激的出现顺序在一场测试以及不同场测试间均应满足随机分布。

听音测试的设计还应保证提供给评价员的信息量不过载，从而避免降低评价员判断的准确性。除非声音和图像间关系十分重要，音频系统主观评价首选无伴随图像的方式。

测试设计中还有一项重要考虑就是采用恰当的控制条件。比如，控制条件包括以评价员不可预测的方式出现的无损音频素材。正是从这些控制刺激与潜在损伤刺激之间的差异判断得出的等级评分才是真正对损伤的评价。

有关测试设计的具体考虑见本标准的后续部分。测试设计、实施和统计分析问题十分复杂，因此，本标准只能给出诸如此类的最普遍的指导方针。建议在听音测试之初咨询实验设计和统计分析相关专家或请专家直接介入。

5 评价小组的选择

5.1 专家评价员

音频系统小损伤听音测试的评价数据应全部来自具备专业小损伤探察技能的专家。待测系统达到的质量越高，越需要专家进行评价。

5.2 评价员的选择

5.2.1 评价员的选择准则

小损伤声音系统的主观测试是基于一组经过选择的评价员，其结果主要不是用于外推到一般大众的听音情况，而是调查一组听测专家，在特定的条件下，能否感知相对微弱的质量下降并给出损伤的定量估计。对测试过程的严格控制是为了揭示被测系统一旦投入使用后，消费者在长期不同条件下的实际生活的使用中可能会发现的问题。

有时，需要在测前或测后使用一种筛除技术（测试之前称为预筛除，测试之后称为后筛除），有时，两种筛除技术都需要。这里，筛除是指一种处理，筛除处理未通过的某个评价员的评价结果将全部被忽略。

任何一种未经仔细分析和应用的筛除技术都有可能导致有偏差的结果，因此，一旦发生数据筛除，报告中应清晰描述所用的筛除准则，以便读者做出自己的判断。

5.2.2 评价员的预筛除

预筛除方法包括：听觉测验、基于以往测试中的经验和表现、基于预测试的统计分析结果。也可通过训练过程进行预筛除。

采用预筛除技术主要是为了提高听音测试的效率，但同时会限制结果的普适性，因此应在两者间进行平衡。

5.2.3 评价员的后筛除

后筛除方法大致分为两类：一类是基于个人评价结果与平均结果的不一致性；另一类是基于评价员做出正确辨别的能力。第一类方法不够科学。若采用本标准规定的方法进行测试，将自动生成第二类后筛除方法所需要的信息。评价员后筛除的统计学考虑参见附录 A。评价员专业技能等级评价方法参见附录 B。

后筛除方法主要用来筛除不能做出正确辨别的评价员。应用后筛除方法需在测试结果中予以阐明。但是，评价员对不同类别的损伤的敏感度是不同的，筛除操作应谨慎。

5.3 评价小组大小

如果总体方差可估计且已知实验精度，则可以预测出评价小组的适当人数（样本容量）。

经验表明，在听音测试条件从技术和操作双方面得到严格控制的情况下，20 位评价员给出的数据通常足以得出适当的结论。如果在测试进程中可进行数据分析，那么当达到了得出适当结论所需的统计显著性，则无需更多评价员的继续参与。

如果被测系统预期达到了近乎透明（几无损伤）的程度，则需要更多的评价员，以保证通过后筛除的评价员的数目足够多。

如果由于某种原因测试条件未得到严格控制，则可能需要更多评价员的参与以达到测试所要求的精度。

评价小组的大小并不是实验精度的唯一制约因素。原则上，按照本标准进行的测试，其结果仅当实际参加测试的评价员为一组听测专家时才严格有效。在此前提下，通过增加评价小组的人数，可能使测试结果因有更多听测专家的参与而显得更具说服力。另外，考虑到存在评价员对不同类型的损伤敏感度不同的可能性，也需要增加评价员的人数。

6 测试方法

6.1 方法概述

本标准采用“带隐藏参考的双盲三刺激”方法，此方法灵敏、稳定，有利于对小损伤的准确探察。

此方法的推荐应用形式也是对损伤探察最为灵敏的形式为：每次只有一位评价员从三个刺激（“A”、“B”、“C”）中自行选择，进行评价。三个刺激中通常将已知参考作为“A”，隐藏参考和被测对象在每个试验中随机分配给“B”和“C”。

测试时要求评价员按照连续五级损伤标度，分别评价“B”和“C”相对于“A”的损伤程度。三个刺激中，“B”和“C”其中之一应与“A”无区别，另一个刺激可能存在损伤，该刺激与“A”相比任何可感知的区别均被解读为损伤。

在这种推荐形式的测试中，评价员一旦给出一个试验的评分就应该能够直接继续到下一个试验，而且在做出评价前片段可重复播放。由此，评价员可以自行控制测试进度。

等级评分标度源于 ITU-R BS.1284 中规定的五级损伤标度，应视为“带支撑点”的连续标度，见表 1。

表1　ITU 五级损伤标度

损伤程度	评分等级
损伤不可觉察	5.0
损伤可察觉，但不至引起不悦	4.0
损伤稍令人不悦	3.0
损伤令人不悦	2.0
损伤令人非常不悦	1.0

注：使用预定义的中间支撑点有可能引入偏差[Poulton,1992]，也可使用不带支撑点描述的数字标度，但必须标示标度的排列方向，这样，可以解决以不同语言进行的对比测试中对支撑点描述的翻译问题。

如果不使用中间支撑点，必须将个体评价员的评价结果根据整体均值和标准偏差进行归一化处理。式（1）可用于在保留原始标度的同时实现归一化。

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} \cdot s_s + x_s \quad \text{.....(1)}$$

式中：

Z_i ——归一化的结果；

x_i ——评价员 i 的评分；

x_{si} ——一场测试 s 中评价员 i 的平均分；

x_s ——一场测试 s 中所有评价员的平均分；

s_s ——一场测试 s 中所有评价员评分的标准差；

s_{si} ——一场测试 s 中评价员 i 评分的标准差。

使用无中间支撑点的评分标度时，不允许以绝对值的形式表示评价结果。

建议评分标度精确到小数点后一位。

测试方法包括两个部分：熟悉或训练阶段和等级评分阶段。

6.2　熟悉或训练阶段

进行正式等级评分之前，评价员应（通过训练）完全熟悉测试工具、测试环境、等级评分过程、评分标度以及评价方法，还应完全熟悉待辨别的损伤。如果进行的是最灵敏形式的测试，评价员应在正式评分之前听过所有的测试素材。在熟悉或训练阶段，建议评价员组成小组（如 3 人一组），便于自由讨论察觉到的损伤情况。

附录 C 给出了一套面向评价员的指导书示例，示例包括“带隐藏参考的双盲三刺激”方法描述等。通过正确的训练，可以将一些具备初级能力的评价员转变为适应测试的专家，并使其在随后的正式等级评分阶段所使用的评分标准趋于稳定。

6.3　等级评分阶段

在当天第一场正式等级评分测试开始时，应向评价员口头介绍主观评价指导书（最好辅以书面材料），还可以展示几个对比听音示例。

中长期的听觉记忆并不可靠，测试过程应完全依赖于短期记忆。在评分阶段，最好采用如附录 C 所述的三刺激系统并结合准瞬时切换的方法，此类切换要求刺激在时间线上严格对齐。

注：如果连续刺激的波形不一致，完全的瞬时切换可能产生人工噪声。因此，推荐使用包括渐弱/变换/渐强过程在内40ms的准瞬时切换时间。

在最严格的测试中，每次只能有一个评价员参与评价，以保障评价员拥有完全的自由度在三个刺激

间随意切换，这种自由度对评价员运用自身判断力充分比较每个试验的刺激间的细微差别是必不可少的。

为减少注意力的分散，评价员应能够在没有视觉引导的条件下切换刺激，如果评价员愿意，他完全可以闭上眼睛以集中精力。切换系统不应产生可闻干扰（如“咔哒”声），否则将严重影响评价员的评价过程。

尽管本标准提倡的评价员自行控制试验速度会导致不同评价员的评分过程耗时不同，但一场等级评分测试不应超过 20 分钟～30 分钟，即一场测试包含不超过 10 次～15 次试验。评价员的疲劳是严重影响其判断准确性的主要因素。为避免评价员疲劳，连续两场测试间的休息时间不得少于一场测试的时间。

7 属性

7.1 概述

本章列出了单声道、双声道立体声、多声道立体声（多至 3/2 声道）和先进声音系统的评估属性。建议“基本音频质量”属性为必选属性，其余为可选属性。

对每个试验，如果要求评价员对一个以上的属性做出评价，有可能给评价员造成应答负担。对一个给定的刺激，如果由于要求评价员回答多个问题而造成了负担过重，将导致对所有评价属性的评价结果的不可靠。

7.2 单声道系统

基本音频质量

唯一的全程属性，用于判断参考与被测对象之间一切可察觉的差别。

7.3 双声道立体声系统

基本音频质量

唯一的全程属性，用于判断参考与被测对象之间一切可察觉的差别。

以下为可选的附加属性：

立体声声像质量

本属性与参考和被测对象在音频事件的声像位置、声像深度感和真实感方面的差别有关。

虽然一些研究发现立体声声像质量有可能受损，但没有充足的研究结果支持将立体声声像质量与基本音频质量的评价独立开来。

注：截止到1993年，大多数双声道立体声系统的小损伤主观评估研究只是使用了基本音频质量属性，而立体声声像质量属性一直是作为一个全程属性隐式或显式地包含于基本音频质量属性中。

7.4 多声道立体声系统

基本音频质量

唯一的全程属性，用于判断参考与被测对象之间一切可察觉的差别。

以下为可选的附加属性：

前方声源声像质量

本属性与前方声源的定位相关，它包括立体声声像质量和清晰度的损失。

环绕声质量

本属性与空间感，环境感，或者特定方向的环绕效果相关。

7.5 先进声音系统

基本音频质量

唯一的全程属性，用于判断参考与被测对象之间一切可察觉的差别。

先进声音系统的属性应包括描述多声道系统的属性。以下为可选的附加属性：

音品--本属性尤为重要

音品属性可通过两组特性描述。

第一组与声音的色彩相关，如，明亮度、音色、着色、清晰度、硬度、均衡度和丰满度。

第二组与声音的均匀度相关，如，稳定度、急剧度、逼真度、保真度和动态。这些性质可以描述音品，也可以描述声音的其他特性。

定位质量

本属性与方向性声源的定位有关，包括立体声声像质量和清晰度的损失。本属性可分为水平定位质量、垂直定位质量和远距离定位质量。在伴随图像的测试中，这些属性还可分为显示器处的定位质量和听音者处的定位质量。

环境声质量—扩展了环绕声质量

本属性与空间感、包容感、环境感、声场扩散性或者空间定向环绕效果相关。本属性可分为水平环境声质量、垂直环境声质量和远距离环境声质量。

8 节目素材

为揭示被测系统间的差异应采用关键性素材。关键性素材是指能够给被测系统造成压力的素材。不存在可用于评估所有被测系统的普适性的节目素材，因此，每次测试均应专门为每个被测系统挑选关键性节目素材。通常，好素材的挑选相当耗时，但除非真正为每一个被测系统找到了关键性素材，否则，就不能揭示被测系统间的差异，测试也不具说服力。

当系统间无差异的检测结果可以被接受为有效之前，必须从经验上和统计上说明，造成不能发现系统间差异的原因不是由于音频素材选取不当或其他薄弱环节引起的测试不灵敏。极端情况下，如果部分或全部系统的测试结果均为全透明，则需要专门设计带低质量或中等质量支撑点（素材）的特殊试验来检验评价员的专业水平（见附录 A）。

支撑素材的质量必须是已知的（如来自以前的研究结果），专家评价员可以察觉而非专业人员无法察觉。支撑素材作为测试条目，不仅可用于对评价人员专业技能的检查，还可以用于对测试环境其他方面灵敏度的检查。

如果采用第六章规定的标准测试方法和附录 A 规定的统计方法，所有评价员均可正确识别隐式地嵌入在近乎透明的条目间或出现于单独安排的测试中的支撑素材，则这可作为评价员具备足够专业能力以及测试环境其他方面也不存在灵敏度缺陷的证据。因此，评价员不能区分编码系统（条目）与未编码系统而得出被测对象近乎透明的测试结果有效，表示的是真正的“透明”。

另一方面，如果评价员未能正确识别出支撑素材，则说明或者是评价员缺少专业技能，或者是测试存在灵敏度缺陷，或者两者兼而有之。因此，“被测系统可以透明传输”的结论缺乏证据，需要替换不能正确识别支撑素材的评价员并进行有助于提高测试灵敏度的其他改进，重新测试。

任何可作为广播电视节目的音频素材均可纳入关键性素材的选取范围，但关键性素材不应包括针对特定系统精心设计的人工信号。为避免分散评价员的专注力，节目素材的艺术性和知识性内容应既不引人入胜也不令人厌烦或乏味。选取关键素材时应考虑实际广播中各种素材类型出现的频率，还应考虑将来广播素材的特性可能随着音乐风格和大众偏好的变化而变化。可利用客观感知模型辅助进行关键性素材的选取。

选择节目素材时，明确待评价的属性很重要。应将素材选择的任务委派给一组对可能出现的损伤有基本认识的有经验的评价员。素材的初选范围应广泛，还可延伸到专用的素材。

为了准备主观比较测试带，每段片段在录制到介质之前都需要由一组专业人员进行主观的响度调整，以保证在后续测试中介质上的所有节目条目均可在固定的增益下重放，为此，该组专业人员应就每个测试片段的相对声级达成共识，还应就序列整体相对于校准电平信号的绝对重放声级达成共识。

每个录音带的开头都应录制一段幅度为校准电平的音频脉冲（例如 1kHz，300ms，-20dB FS），用于将其输出的校准电平校准至重放声道要求的输入校准电平（见 10.4.1）。采用数字方式录制时，校准电平应为-20dB FS，录制时，还应控制节目信号的峰值幅度，使其不超过 GY/T 282—2014 规定的节目最大真峰值电平最大值（-2dB TP）。录制音频脉冲也有利于参考刺激和测试刺激的时间对齐。

测试片段的数目取决于具体测试，其最小值为 5，合理估算值为被测对象数目的 1.5 倍，一项测试中的片段数对每一被测对象应相同。音频片段的典型时长为 10 秒至 25 秒。挑选关键性素材任务复杂，应合理制定时间表，且保证被测对象随时可用。

对于单声道和立体声系统的评估，如果片段选自易于获得的音频源（如 SQAM 激光唱片），则更方便随时进行测试带素材与原始素材的对比检查。但是，相比于音频源是否容易获得，使用真正的关键性素材才是更为重要的。

在双声道重放条件下多声道系统的性能测试应使用参考下混系数。使用固定下混系数，尽管有时会起到限制作用，但从长远来看，无疑是广播业者最明智的选择。参考下混等式（式（2），来自建议书 ITU-R BS.775）为：

$$\begin{cases} L_0=1.00L+0.71C+0.71L_S \\ R_0=1.00R+0.71C+0.71R_S \end{cases} \quad \cdots\cdots (2)$$

测试先进声音系统时，先进声音系统下混至双声道或多声道系统所用的下混等式，或者是将音频对象提供至具体声道的执行过程，应在测试报告中加以描述。

对使用参考下混等式生成的双声道性能进行评估时，测试片段的预选应基于下混生成的双声道节目素材的重放。

9 重放设备

9.1 概要

基准监听扬声器或耳机的选择目标为：所有声音节目信号或其他测试信号都能以最佳效果重放。也就是说，对任何重放都应产生中性的声音，并且可用于单声道，双声道和多声道立体声系统的评价。

某些声音质量缺陷在使用耳机重放时更容易被感知，另一些在使用监听扬声器时更容易被感知。因此，有必要通过主观预测试确定适当的重放设备类型。特别是是当质量缺陷表现在立体声声像特性时，应使用扬声器重放。

评价双声道立体声系统时，使用立体声扬声器和耳机；评价单声道系统时，使用一个中置扬声器和（或）耳机。

对单个试验或成组试验来说，如果监听扬声器与耳机只能选择其一，那么听音效果可能受到换能器的影响而使评价员的有效人数减少。如果评价员可以在监听扬声器与耳机间随意切换，则可免受换能器的影响。

对于伴随和不伴随图像的多声道声音系统和先进声音系统，如果待评价项为所有声道同时重放时的影响，则应使用扬声器。

在相关频率范围内，所用每个扬声器都必须在声学特性上相匹配，以保证彼此间的固有音色差最小。

9.2 基准监听扬声器

9.2.1 概要

基准监听扬声器是指高质量的演播室监听设备，由放置在特定尺寸机箱内的配备专业均衡器的扬声器系统集成单元、高质量功放和分频网络组成。

监听扬声器电声特性必须满足如下最低要求（测量于自由场）。如无特殊规定，绝对声级在距离声中心 1 米处测量。

注：声中心是用于测量用途的参考点，通常对应扬声器最高频率辐射面的几何中心点，应由生产商标明。

9.2.2 电声要求

9.2.2.1 幅频响应

以粉红噪声为信号源，在 40Hz～16kHz 范围内，主声轴方向（方向角为 0°）处每三分之一倍频带上测量的幅频响应容限应为一个不超过 4dB 的通带；±10°方向角（与主声轴夹角）处测量的频响曲线与主声轴方向的差值（仅限同一水平面内）不应超过 3dB；±30°方向角的差值不应超过 4dB。

不同扬声器的幅频响应应一致，至少在 250Hz～2kHz 的频率范围内，不同扬声器频响的差值不应超过 1.0dB。

9.2.2.2 指向性指数

以三分之一倍频带噪声测量的指向性指数 C，在 500Hz～10kHz 频率范围内应满足：

$$6\text{dB} \leq C \leq 12\text{dB}$$

指向性指数应随着频率平滑上升。

9.2.2.3 非线性失真

以产生 90dB 平均声级（SPL）的恒定电压信号加载到扬声器，相对于该声级，在 40Hz～16kHz 基频范围内，谐波失真分量不应超过：

-30dB（3%）　　f ＜250Hz

-40dB（1%）　　f ≥250Hz

9.2.2.4 瞬时保真度

主声轴方向，以示波器测量的信号衰减到初始幅度 1/e(大约 0.37)的衰变时间应满足：

$$t_s < 5 / f$$

其中 f 为信号频率。

即正弦猝发音的衰变时间不超过相应正弦波周期的五倍。

9.2.2.5 时延

立体声或多声道系统的不同声道时延差应不超过 100μs。

注：其中不包括从扬声器到听音位置的时延。

对伴随图像的系统，基准监听扬声器和被测系统的总时延差不应超过 ITU-R BS.775 中的限定值。

9.2.2.6 动态范围

在满足至少连续工作 10 分钟不产生任何热损伤，机械损伤，电路不过载的情况下，监听扬声器产

生的最大工作声级（以符合 GB/T 6278—2012 的节目噪声信号测量）应满足：

$$L_{\text{eff max}} > 108\text{dB SPL}$$

$L_{\text{eff max}}$ 为声级计在不计权、慢时间常数下的均方根测量值。

单只基准监听扬声器和相应放大器在距声中心 1m 处产生的等效噪声声级应满足：

$$L_{\text{noise}} < 10\text{dBA}$$

9.3 基准监听耳机

9.3.1 概要

基准监听耳机是指高质量的演播室监听设备，监听效果类似于扩散场响应。

9.3.2 电声要求

9.3.2.1 频率响应

演播室监听耳机的扩散场频率响应见 ITU-R BS.708。

9.3.2.2 时延

立体声系统不同声道间的时延差不应超过 20μs。

在伴随图像的系统中，基准监听耳机和被测系统的总时延不应超过 ITU-R BS.775 中的限定值。

10 听音条件

10.1 概要

“听音条件”描述了对采用扬声器重放时产生的基准声场的复杂的声学要求，该声场直接影响评价员在基准听音位置的听音。听音条件导致了听音位置或听音区的声场特性，具体包括：

——听音室的声学特性；

——听音室内扬声器的安排；

——基准听音点或听音区的位置。

鉴于目前的技术发展水平，仅用声学参数尚不足以完整且唯一地描述基准声场特性，因此，本标准还列出了对听音室的几何特性和室内声学要求。

10.2 基准听音室

10.2.1 概要

使用扬声器重放进行主观测试时，基准听音室应满足如下最低要求。

仅使用耳机重放时，听音室至少应满足对背景噪声声级的要求。

10.2.2 描述基准听音室合适的净尺寸。如果听音室不满足尺寸要求，那么至少应满足声场条件要求和对扬声器的安排要求。

10.2.2 几何特性

10.2.2.1 房间大小（地面面积）

房间大小应满足如下要求：

——单声道或双声道立体声重放：20㎡～60㎡。

——多声道立体声或先进声音系统重放：30㎡～70㎡。

注1：房间过小将限制同时参与听音的评价员的最大人数。

注2：先进声音系统听音室的最佳特性还有待于进一步研究。房间的大小，形状，比例和声学特性应在测试报告中写明。

10.2.2.2 房间形状

房间应以前置立体声扬声器底座中心点连线的中垂线平面为轴对称分布，地面区域最好为矩形或梯形。

10.2.2.3 房间比例

房间长宽高的比例应满足以下要求以保证房间低频本征频率的合理均匀分布：

$$1.1\,w/h \le l/h \le 4.5\,w/h - 4$$

其中：

l ——长；

w——宽；

h ——高。

另外，还应满足 $l/h < 3$ 和 $w/h < 3$。

10.2.3 室内声学特性

10.2.3.1 混响时间

在 200Hz～4kHz 的频率范围内，混响时间均值 T_m 应满足式（3）的要求。

$$T_m = 0.25\,(v/v_0)^{1/3} \quad \text{.....(3)}$$

式中：

T_m——混响时间均值，单位为秒（s）；

V ——室内容积；

v_0——参考容积 100㎥。

频率范围 63Hz～8kHz 的 T_m 的容差范围见图 2。

注：对低频处的短混响时间，很难测量。

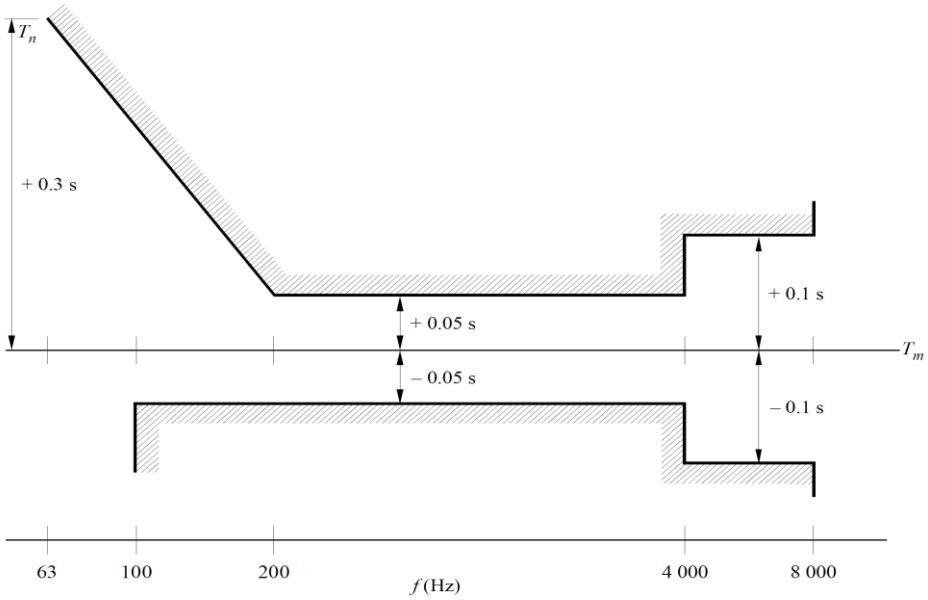

图 2　混响时间相对于平均值 T_m 的容差范围

10.3　基准声场条件

10.3.1　概要

听音区的声场特性，对于评价员的主观感觉、声音质量的评价、不同听音地点或房间进行的评价结果的再现性至关重要。声场特性源于扬声器和听音室的相互作用，并与所使用的听音安排（见 10.5）有关。

10.3.2　直达声

10.3.2.1　监听扬声器频率响应

监听扬声器在自由场下的频率响应应满足 9.2.2 的要求。

10.3.3　反射声

10.3.3.1　早期反射

由听音室界面反射引起的、在直达声到达听音区后 15ms 内到达听音区的早期反射声，在 1kHz～8kHz 的频率范围内相对于直达声应至少衰减 10dB。

10.3.3.2　后期能量

除对早期反射和混响的要求外（见 10.2.3），还应避免其他显著的声场异常，例如颤动回声，声染色等。

10.3.3.3　混响时间

见 10.2.3.1。

10.3.3.4　脉冲响应

测试报告中应描述听音室在正常使用状态下（包括装修、陈设）从每个扬声器发出的，在所有可能的听音位置处的脉冲响应，以时域表示。这有助于确认扬声器与房间声场一起在早期反射声，后期能量和混响时间特性上满足要求的程度。

10.3.4 稳态声场

10.3.4.1 听音室响应曲线

听音室响应曲线定义为，以 50Hz～16kHz 的粉红噪声为信号源，由每个监听扬声器在基准听音位置产生的三分之一倍频带声级频率响应。听音室频率响应曲线的容差范围见图 3，图中 L_m 为声级均值。

听音室中各扬声器在基准听音位置产生的房间响应曲线之间的差异在整个频率范围内不宜超过 2dB 的容差。前置扬声器（±60°方位角），尤其是房间中部高度那一层的前置扬声器的性能匹配最为重要，报告中应列出听音室房间响应的测量曲线。允许使用均衡以满足本指标要求，但应在报告中列明使用了均衡以及所使用的均衡的细节。

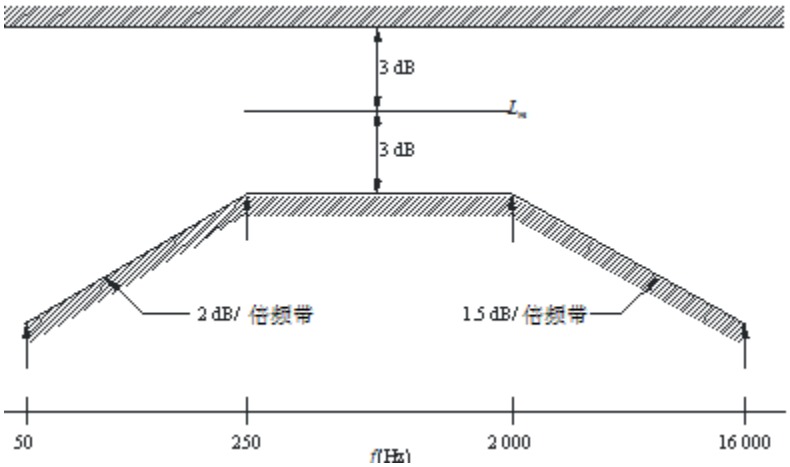

图 3 听音室响应曲线容差范围

10.3.4.2 背景噪声

在听音区听音者落座后耳部的标称高度处测量的连续背景噪声（源于空调系统，内部设备和其他外部声源）不宜超过 NR 10。

任何情况下背景噪声都不应超过 NR 15。

1/3 倍频带噪声评价曲线 NR 10（推荐）和 NR 15（最大）见图 4，倍频带噪声评价曲线 NR 10（推荐）和 NR 15（最大）见图 5。

背景噪声不应为周期性或音调性的可感知脉冲。

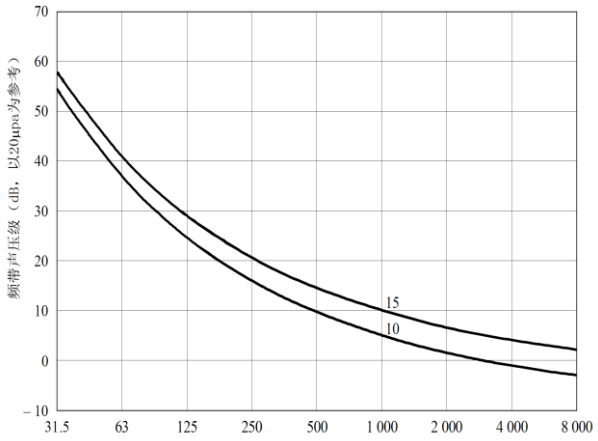

图 4　1/3 倍频带背景噪声评价曲线 NR 10（推荐）和 NR 15（最大），基于 ISO R1996（1972）

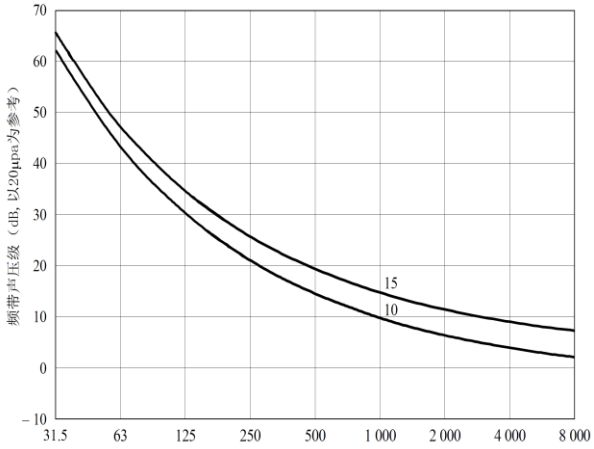

图 5　倍频带背景噪声评价曲线 NR 10（推荐）和 NR 15（最大），基于 ISO R1996（1972）

10.4　听音声级

10.4.1　扬声器重放

10.4.1.1　工作声级（基准听音声级）

基准听音声级定义为：重放给定的测量信号，在基准听音位置产生的理想听音声级。该指标描述了重放声道的声增益，以确保相同片段在不同听音室的重放声级相同。

应使用粉红噪声对听音安排中的每一扬声器进行声级校准。

将幅度（均方根电平）为校准电平的测量信号（按照标准 ITU-R BS.645，幅度为 0dBu0s；按照 GY/T 192—2003，为低于数字磁带录音的削波电平 20dB），依次馈入每个重放声道（如功放及其扬声器）的输入端，调整放大器的增益至产生基准声级（IEC/A 加权，慢档）。

$$L_{ref} = 78 \pm 0.25\ (\text{dBA})$$

（测试经验表明，个别评价员喜欢采用不同的绝对听音声级，本标准对这样的特殊需求不推荐也不阻止。听音声级不同是否会影响某些待评价损伤的可闻度在现阶段还不能确定，因此，如果评价员调整了系统增益，需要在测试结果中加以标注。）

10.4.1.2　耳机重放

调整耳机增益，使评价员感觉到的响度与采用扬声器重放时在基准听音位置上感觉到的基准声场响度一致。

10.5　听音安排

10.5.1　概要

听音安排描述听音室中扬声器和听音位置（听音区）的设置情况。

听音测试通常都在基准听音位置和其他推荐的位置进行。然而，明显偏离听音中心造成的影响也有必要加以评估。因此，本标准中也给出了“最差”听音位置。

10.5.1.1 监听扬声器的高度和方向

与评价员在同一高度层上的所有监听扬声器的高度（测量至扬声器的声中心）均应为评价员就座时耳部的高度。扬声器的方向应保证其参考声轴通过基准位置上评价员就座时的耳部高度处。如果先进声音系统包括置于不同高度层的扬声器，则有必要描述所有扬声器在水平和垂直方向上的相对于房间大小和听音点的位置。

10.5.1.2 距墙距离

对于非内嵌式的扬声器，扬声器的声中心距周围反射面至少应 1m。如果受房间尺寸的限制难以达到，应使用其他方法控制早期反射声以满足 10.3.3.1 的要求，并应在测试报告中陈述不满足与墙的距离要求及控制早期反射声所使用的方法。

10.5.2 单声道重放

对只使用一个扬声器的单声道信号重放，最小听音距离应为 2m，所有的听音位置均应在扬声器轴线±30°范围内。使用扬声器 M 的单声道音频系统主观评价基准听音安排和允许的听音区见图 6，图中 D 为听音距离，表示在同一高度上，扬声器和听音者之间的直线距离。

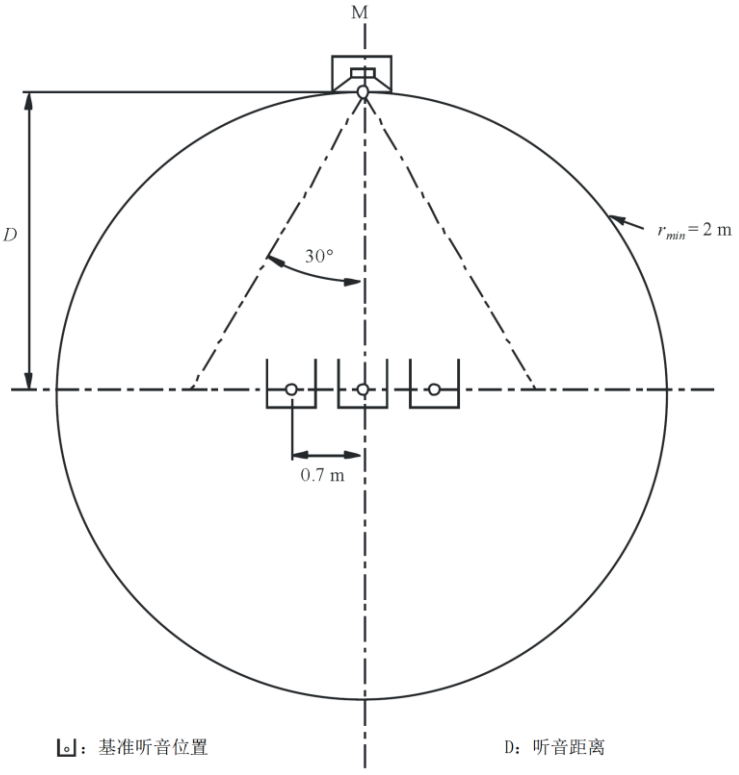

图 6 使用扬声器 M 的单声道音频系统主观评价基准听音安排和允许的听音区

10.5.3 双声道立体声重放

10.5.3.1 底座中心间距（B）

两个扬声器底座中心间的距离 B 的推荐值为 2m～3m。对合理设计的房间，B 值达到 4m 也是可接受的。

10.5.3.2 听音距离（D）

在同一高度上，扬声器和听音者之间的直线距离 D 为（2～1.7B）m。

10.5.3.3 听音位置

使用扬声器 L/R 的小损伤立体声音频系统主观评价听音安排见图 7。

基准听音位置定义为如图 7 所示的听音距离 D=B 时，60°听音角的顶点。基准听音位置和“最差”听音位置见图 7。

推荐的听音区域是围绕基准听音位置，半径不超过 0.7m 的区域。

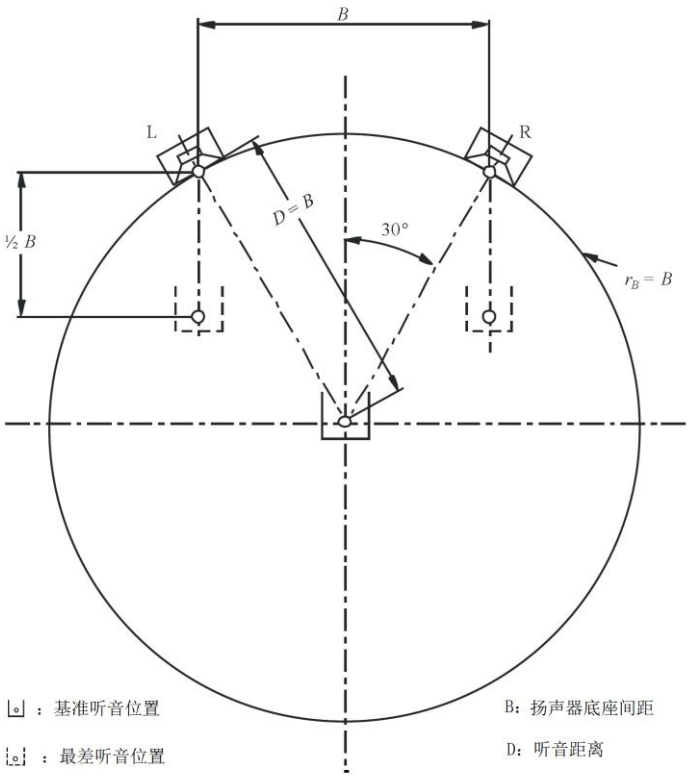

图 7 使用扬声器 L/R 的立体声音频系统小损伤主观评价听音安排

10.5.4 多声道立体声重放

听音安排原则上应符合 ITU-R BS.775 图 1“使用扬声器 L/C/R 和 LS/RS 的基准扬声器安排”中所列的 3/2 多声道布局。

10.5.4.1 底座中心间距

L 和 R 两个扬声器底座中心间的距离 B 的推荐值为 2m～3m。对合理设计的房间，B 值达到 5m 也是可接受的。

10.5.4.2 听音距离和基准角

基准听音距离为 B，基准角等于 60° 。

10.5.4.3 听音位置

使用扬声器 L/C/R/LS/RS 的小损伤多声道音频系统主观评价听音安排见图 8。

基准听音位置定义为如图 8 所示的 60° 听音角的顶点。基准听音位置和“最差”听音位置见图 8。

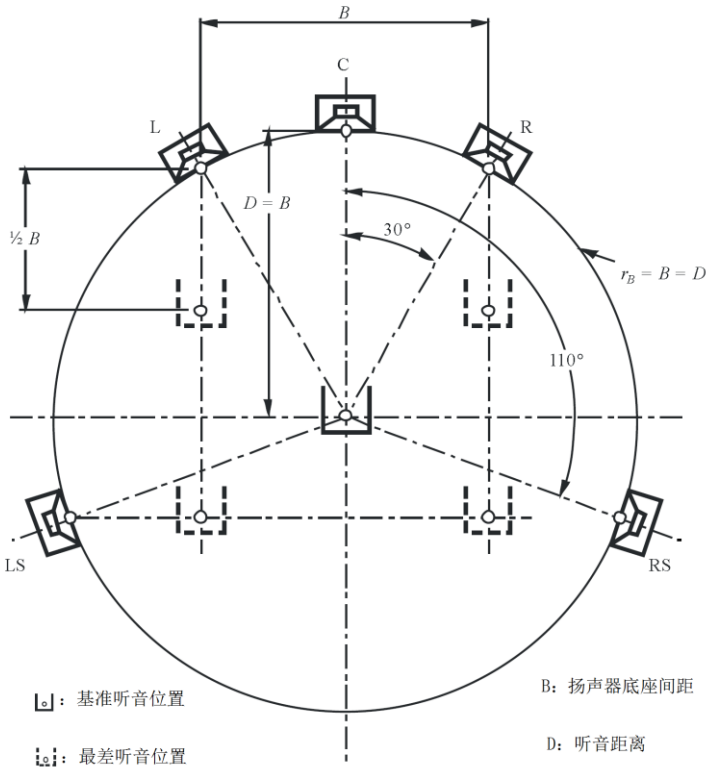

图 8 使用扬声器 L/C/R/LS/RS 的多声道音频系统小损伤主观评价听音安排

10.5.5 先进声音系统的重放

测试报告中应详述所用扬声器的位置（距离和角度），以及相对于听音位置的定位，以阐明测试条件。应按照 ITU-R BS.775 中对相应细节内容的描述方式记录扬声器布局和听音位置。对使用处于不同高度的扬声器的先进声音系统，测试报告中还需在垂直维度上描述扬声器的位置。具体细节信息见 ITU-R BS.2051。

11 统计分析

测试结果统计分析的根本目的是精确确定每一个被测系统的平均性能以及代表这些平均性能的数字间差异的可靠性。后者要求进行结果的变化性或方差估计。

按照本标准，评分标度是区段化的，各区段间隔相等。评分结果的等级属性对统计方法没有特殊要求，既不禁止也不规定采用何种统计方法。

如果参数统计基于的假设合理，则该方法灵敏有效，推荐使用。仅当数据的重要属性与 ANOVA 基于的假设严重背离时，才应该重新选择分析方法（如非参数的方法）。具体来说，推荐首先使用 ANOVA 作为第一阶段的基础分析方法，然后，借助 ANOVA 方差估计的结果，利用其他方法（例如 t 检验，Neuman-Keuls，Scheffe，等）研究细节信息，进一步研究 ANOVA 发现的显著效应。

某项假设经常可通过多种统计方法来验证，如果存在其他的统计方法也可以验证某一假设的有效性，那么结论将更有说服力，因此建议进行辅助数据分析（例如 Wilcoxon 等）。

在某些阶段需要考虑心理学方面的影响，这无疑会对从非物理尺度得出何种有意义的结论产生影响。

需要注意的是，除非评价尺度是线性的，否则不同等级的比较只能基于等级次序。

12 统计分析结果陈述

12.1 概要

统计分析结果的陈述应能使专业和非专业人员都能从中得到相关信息。首先，读者希望看到整体的测试结果，使用图表的形式更一目了然。结果的陈述还需要更详细的定量信息，应在报告的附录中给出完整详细的数据分析。

12.2 绝对评分值

分别给出被测对象和隐藏参考的绝对评分均值，有利于读者初步了解数据的概况。

但绝对评分值不宜作为详细统计分析的基础，这是因为使用本标准规定的测试方法时，评价员明确知道在成对出现的刺激中，有一个隐藏参考与源是一致的，因此，观察样本并不独立，对绝对评分值的统计分析也就无法得出有意义的结论。

12.3 评分差值

隐藏参考与被测对象之间的评分差可用于进一步的统计分析。图形式的表示可以清晰看出读者首要关注的被测对象质量离达到透明的距离。

12.4 显著性水平和置信区间

测试报告应向读者提供所有主观数据的固有统计特性信息。报告中应注明显著性水平以及有助于读者理解的有关统计方法和结果的其他细节信息，例如以图形表示的置信区间或误差线。

“正确的”显著性水平是不存在的，按惯例，显著性水平一般取 0.05。原则上，根据被测假设选择使用单侧或双侧检验。

13 测试报告内容

测试报告应尽可能清晰地表述测试理由、测试方法和测试结论。报告应提供详尽信息，原则上，足以使专业人士重现测试并实证结果；使有经验的读者能够理解和评判测试的主要细节，如测试目的，测试的设计、执行过程，以及分析和结论。

应特别注意以下几点：

——评价员和片段的说明和选择；

——听音环境和设备的物理细节，包括听音室尺寸和声学特性，重放设备的类型和摆放，电子设备的规格；

——声道布局是否符合 ITU-R BS.775 或 ITU-R BS.2051 的描述；

如果声道布局不符合 ITU-R BS.775，则应按照 ITU-R BS.775 中对相应细节内容的描述方式记录所有的扬声器位置，以允许再测时的位置还原，还应记录相对于扬声器位置的基准听音位置（见 10.5.4 和 10.5.5）；

——是否满足 10.5.1.2 的距离要求，如不满足，需记录，并记录控制早期反射声使其满足 10.3.3.1 要求所使用的方法；

——所测得的所有扬声器在听音室产生的房间响应，如果使用了均衡处理，应承认并记录使用的均

衡方法；

——应报告听音室声学和物理特性与本标准的偏差，包括：10.3 中规定的听音室的声学测量结果、10.4 规定的听音电平和 10.5 要求的物理距离；

——听音室在正常使用状态下（包括装修、陈设）从每个扬声器发出的，在听音位置处测量的房间的脉冲响应，以时域显示；

——测试设计，训练，给评价员的说明，测试序列，测试过程，数据生成；

——数据处理，包括描述性的和分析性推理统计的细节；

——得出结论的基础。

附 录 A
（资料性附录）
评价员后筛除的统计学考虑

A.1 评价员专业技能的评估

评价员按照“带隐藏参考的双盲三刺激”方法逐个进行评价，对每个试验给出两个等级评分，评价时，可以直接对比一个试验的两个评分，也可以通过所有试验检查个人的评分值。统计时，取一个试验的两个评分的代数差，代数差的计算方向对所有序列都相同，假定总是用被测对象减去隐藏参考的评分。

在一次听音测试中，如果评价员未能从总体上正确识别出隐藏参考与被测对象，则该评价员给出的所有序列评分差中会出现正值，计算评分差均值时会出现正负抵消的情况，从而使评分差均值为 0 或接近于 0；相反，如果评价员总体上正确识别出隐藏参考与被测对象，那么大部分评分差为负值，从而评分差的均值亦为负值。

这样，形成的数据服从单边 t 检验，检验每个评价员评分差分布均值为 0 的可能性。如果对某一评价员，零假设被拒绝，则可以断定，在给定置信度下，该评价员的数据源自均值小于 0 的分布。也就是说，对于服从均值小于零的分布的每一位评价员，已经证明，他（她）总体上并不是靠猜测做出正确判断，而是真正具有足够的专业技能，完全可以将该评价员的数据纳入测试结果的统计之中。反之，则可以断定该评价员的数据是依靠猜测的，不能参与后续的统计分析。

本标准仅适用于小损伤评价。无论由于何种原因，如果测试中不仅仅包括“小”损伤，还包括了很多“大”损伤，则简单应用后筛除的方法可能导致虚假的或不当的结论。这里，“大”损伤是指即使对于非专业评价员也是易于察觉的损伤。很明显，隐藏在多数“大”损伤中的少数真正的“小”损伤，在 t 检验中所占的权重很小。t 检验中，t 幅值的最大权值出现在大损伤条目，从而使小损伤条目的分差值迷失在统计噪声里。于是，确实能够正确判定小损伤条目的专家评价员将湮没在靠猜测评分的评价员中，无法区分。

即使在严格的小损伤评价中，也难免出现大损伤条目，但只占评价条目的少数。鉴于此，为使后筛除足够严格，通常应在评估评价员技能的 t 检验中排除大损伤条目。大损伤条目的评分均值很低，比如，分差均值在-2.0 至-4.0 之间。大多数评价员都能够正确区分其隐藏参考与被测对象，因此，在 t 检验中包含大损伤条目，对评估评价员的专业水平是弊大于利，将导致夸大或过高估计评价员的专业技能。

反之，测试中包含很多“真正透明”的条目，在 t 检验中也宜忽略这些近乎透明（“太难”）的条目的数据，以使一些预计有影响的条目在测试中更有分量，否则会低估评价员的专业水平。

通常，“过难”或“过易”的条目同样都不适合用于区分专家和非专家评价员。

正确操作的测后 t 检验的独特的优点是，评价员在一项测试中是否具有足够的专业水平可以通过其在该项测试中的评价数据得以评估。在涉及相同评价员的一系列测试中，可能会发现，如果一名评价员顺利通过了预筛除，并不一定能通过后筛除，部分评价员只能通过这一系列测试的一个子集，需要根据具体结果来判定在后续的统计中保留或删除某一评价员的数据。实际上，后筛除方法在预筛除特性的基础上，还体现了对“专业水平”这一概念微调的功能。

专业水平不足的评价员不能给出有效数据，因此，使用严格的后筛除的方法来客观确定是否彻底删除专业水平不足的评价员数据是合理的。另一方面，也无法保证通过 t 检验后筛除的评价员给出的数据就一定是好数据。举一个极端的例子，一位评价员完全正确区分出了所有试验的隐藏参考和被测对象，但对所有被测对象的评分均为 1.0，换句话说，来自该评价员的数据集均由-4.0 的评分差构成。

假设在该项测试中，其他评价员给出的分值数据都呈正常分布，只有该评价员的数据比较奇怪（分

差均为－4.0），那么可能会出现删除该评价员数据的主张。然而，除了此处为了说明之便列举的比较极端的例子外，事实上，很难用这种“后验”的方法来确定是否接受数据。这相当于故意按照实验人员的预想塑造数据，而拒绝接受实验的实际结果。

不应使用此类“后验”的方法。只要评价员的总数足够大，即使是明显偏离的个别数据对数据集造成的影响也会很小，因此，测试结果通常是有效且可重现的。测试结束后，如果对试验数据有所怀疑，唯一的办法是使用完全不同的评价人员重新开始整个测试，并努力纠正以前测试过程中的可疑纰漏。

A.2 评价员专业技能的进一步评价

随着基于感知的有损编解码质量的不断提高，具有充足专业技能可以察觉编码仍存损伤的评价员数目必将越来越少。由于容易察觉的损伤减少，对于以前包含相对易察觉的损伤的测试来说具备充足专业技能的的评价员，现在就未必称职。而且，尽管一个评价员的 t 值能够说明其对测试整体具有足够的专业技能，但仍旧有可能无法区分个别质量很高的编码信号和参考信号。这种情况下，评价员的数据就会引入“统计噪声”，掩蔽其他评价员真实感觉到的损伤。

附 录 B
（资料性附录）
评价员专业技能等级评价

通常，给定测试中一位评价员的所有数据用于评估其 t 值。具有足够高 t 值的评价员的数据用于ANOVA。

建议对评价员给出的数据子集反复进行多次 t 检验，检验标准的严格性逐次递增。

如果通过再次评估证明某评价员具有足够的专业技能，那么将其数据用于随后的 ANOVA。反复评估的次数越多，通过评估的评价员的专业技能越强。评价专业技能的标准如下所示。

使用假设数据集，t 检验前舍弃数据点的方法见图 B.1。首先，计算某评价员所给数据的均值和标准差，以用于确定该评价员数据相应的 z 值。z 值代表标准化为服从均值为 0，标准差为 1 分布的值，定义为$z = \frac{x - \mu}{s}$，其中 x 为数据点，μ为样本均值，s 为标准差，s 的计算公式见式（B.1）。

$$s = \sqrt{\frac{N\sum x^2 - \left(\sum x\right)^2}{N(N-1)}} \quad \text{（B.1）}$$

z 值超过标准（μ+Δ1s）的数据点一律排除，用剩余的数据点进行再次 t 检验。图 B.1 中，超过μ+Δ1s 的数据点（阴影区）被舍弃，剩余数据（非阴影区）进行后续的 t 检验。如果在舍弃数据点后进行的 t-检验中，评价员仍显示出足够的专业技能，那么该评价员的所有的数据都将用于后续的 ANOVA。否则，该评价员的所有数据都不出现在后续的 ANOVA。接着用更严格的参数μ+Δ2s 重复删除数据点过程，直至用μ+Δis 重复 N 次，i=0,1,..,N。Δis 和 N 的适当取值，相关机构正在研究中。

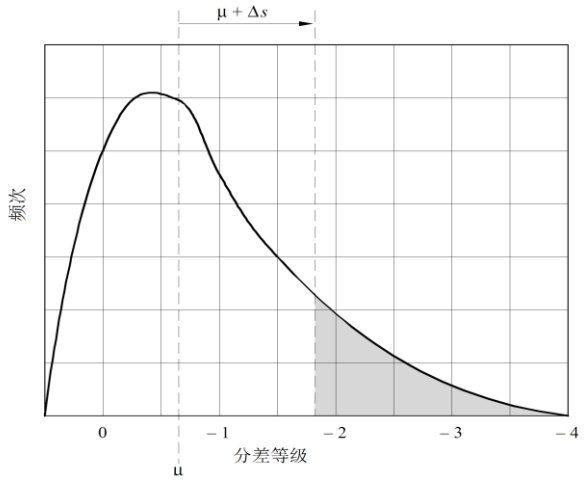

图 B.1 t 检验前舍弃数据点的方法

附 录 C
（资料性附录）
给评价员的主观评价指导书范例

指导书中使用的术语未严格遵循本标准第3章的定义。

C.1 熟悉或训练阶段

训练阶段的目的是使评价员识别和熟悉被测系统可能产生的失真和损伤。训练结束后，您（受训者）应明确“要听的是什么”。今天上午您试听的音频素材将用于下午的盲测。训练阶段，您还将熟悉测试的程序。

您将听到音频素材的参考（原始）版本和经过处理的版本。在视频监视屏幕上，参考版本标识为“A”，信号经过处理的版本和隐藏参考标识为“B”和“C”。使用时，您可以在“A”、“B”或“C”中随意切换，以进行三者之间精细的比较，并根据“A”与“B”以及“A”与“C”之间的区别的大小给出等级评分。音频序列时长通常为10秒～25秒，可根据您的需要无限次重放，您最多有三个小时的训练时间熟悉下午即将正式盲测和评分的所有素材。训练阶段您可以随意使用扬声器或耳机。

测试中，您需要根据表C.1所列的评分标度进行等级评分。

表C.1 评分标度

损伤	等级评分
觉察不到损伤	5.0
损伤可觉察，但不令人不悦	4.0
损伤稍令人不悦	3.0
损伤令人不悦	2.0
损伤令人非常不悦	1.0

（向评价员解释评分标度的含义。应强调：等级评分标度可视为定义了特殊支撑点的连续等间隔标度。）

下午的测试中，每个试验都包含一个隐藏参考（参考素材的完美复制），因此，每个试验中至少有一个等级评分为5.0（且仅有一个，这里建议每个试验仅有一个5.0是出于如下考虑：有些评价员原本可以察觉非常细微的损伤，但为避免犯错误，保守起见，给了两个5.0分，因此本项规定的目的是强制评价员表明哪一个素材是经过编解码的）。如果您发现“B”或“C”的质量优于参考，这意味着差别为“损伤可觉察，但不令人不悦”，可根据所感知的差别，给4.0到4.9之间的评分。

训练过程中，您应考虑，作为一个个体，如何根据评分标度解释所听到损伤，且在任何时候均不应与其他评价员讨论个人的解释。

C.2 训练内容范例

第一天上午持续3小时的主要训练可以成组进行，每组4名评价员。每个评价员均应提前收到一份书面指导书。

训练阶段应包括：

——对测试的宗旨和目标的简短介绍；
——所选测试片段的重放，使评价员熟悉即将评价的节目素材及其声音表现形式；
——被测系统的简短解释，素材预选小组确定的损伤种类的口头描述；
——用一些严重受损的条目演示损伤；
——待评价的属性的解释；
——五级损伤标度的解释；
——素材切换和评分训练。
后续进行正式测试之前，应提醒评价员训练阶段的要点内容，还可以包括重听测试素材。

C.3 盲评阶段

盲测的目的是对上午训练阶段听到的各种素材给出评分。

每个试验，您将听到给定音频素材的三个版本，在视频监视器屏幕上分别标注为“A”、“B”和“C”。“A”固定为参考，您需要将“B”和“C”与“A”对比并给出评分。“B”和“C”其中之一为经处理的版本，另一个是隐藏参考（与源相同），但您并未被告知哪个是经处理的版本，哪个是隐藏参考，因此称为盲评。您可以在“A”、“B”和“C”间随意切换，重复聆听，直至确信地给出评价，之后，您可以自行进入下一个试验。

在每个试验中，您需要根据表 C.1 中的五级评价标度对“B”与“A”以及“C”与“A”之间可察觉区别给出评分。因此，每个试验须为“B”和“C”各给出 1 个评分。每个试验至少有且仅有一个 5.0，请在每个试验结束时将评价结果输入计算机。

除了采用将结果输入计算机的方式，也可以使用纸质评分表。

整个盲评阶段都提供表 C.1 的副本。

需要向评价员介绍评分标度的含义，并强调等级评分标度可视为在特殊数值处定义了支撑点的连续等间隔标度。

参 考 文 献

[1] ITU-R BT.500-13 Methodology for the subjective assessment of the quality of television pictures.

[2] ITU-R BT.710-4 Subjective assessment methods for image quality in high-definition television.

[3] ITU-R BT.811-1 The subjective assessment of enhanced PAL and SECAM systems.

[4] POULTON, E.C. [1992] Bias in quantifying judgments. Lawrence Erlbaum Associates, Hillsdale, United States of America, 1992.

中　华　人　民　共　和　国
广播电影电视行业标准
音频系统小损伤主观评价方法
GY/T 298－2016
*
国家新闻出版广电总局广播电视规划院出版发行
责任编辑：王佳梅
查询网址：www.abp2003.cn
北京复兴门外大街二号
联系电话：（010）86093424　　86092923
邮政编码：100866