

ITU-R BS.1770-5 建议书

(11/2023)

BS系列：广播业务（声音）

**测量音频节目响度和真正峰值
音频电平的算法**



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）
BT	广播业务（电视）
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 BS.1770-5建议书

测量音频节目响度和真正峰值音频电平的算法

(ITU-R 135-2/6号研究课题)

(2006-2007-2011-2012-2015-2023年)

范围

本建议书规定了用于确定主观的节目响度和真正峰值信号电平的音频测量算法。

关键词

3/2多声道音响系统，高级音响系统，响度，真正峰值信号电平

国际电联无线电通信全会，

考虑到

- a) 现代的数字声音传输技术提供了一个极宽的动态范围；
- b) 现代数字声音产生和传输技术提供了ITU-R BS.775建议书中详述的一种单声道、立体声和3/2多声道格式的混合体以及在ITU-R BS.2051建议书中详述的高级音频格式，包括基于声道的、基于对象的和基于场景的输入信号及其与元数据的组合，声音节目是采用所有这些格式制作的；
- c) 收听者要求对于不同的声音来源和节目类型，音频节目的主观响度是相同的；
- d) 有许多方法能够测量音频电平，但是在节目制作中采用的、现有的测量方法不能够提供主观响度的指示；
- e) 在交换节目的响度控制方面，为减少对观众的打扰，必需有一个推荐的算法用于主观响度的客观评估；
- f) 将来基于心理声学模型的复杂算法可能为各种各样的音频节目提供改进的响度客观测量；
- g) 数字媒体突然地过载，因而甚至应该忽略瞬时过载，

进一步考虑到

- a) 由于通常采取的处理例如滤波或者比特率降低，峰值信号电平可能会提高；
- b) 由于真正峰值可能出现在两个样本之间，现有的计量技术不能够反映出包含在数字信号中的真正峰值电平；
- c) 数字信号处理的状况使实现一个算法精确地估计信号的真正峰值电平变得切实可行；
- d) 真正峰值电平指示算法的使用将允许准确地指示在数字音频信号的峰值电平和限幅电平之间的净高，

建议

- 1 当要求音频通道或者节目（按照ITU-R BS.775建议书生成到五个主要通路（单源、立体声以及3/2多声道音频））响度的客观测量要便于节目传送和交换时，应采用附件1中规定的算法；
- 2 当要求对一个音频节目的响度进行客观测量（生成的音频节目有很多通路（例如ITU-R BS.2051建议书中规定的信道配置））时，应当使用附件3中规定的算法；
- 3 当需要客观测量基于对象的音频信号或基于声道和对象的音频信号的组合的响度时，应使用附件4中规定的算法；
- 4 在节目制作和后期制作中采用的、用于指示节目响度的方法可以以附件1、3和4中规定的算法为基础；
- 5 当需要指示数字音频信号的真正峰值电平时，测量方法应基于附件2中指出的准则，或者一个能提供类似或较好结果的方法，

进一步建议

- 1 应该考虑到如果证实新的响度算法提供的性能要明显优于附件1、3和4中规定的算法，可能需要更新本建议书；
- 2 当开发出新的算法时，应当对该建议书进行更新，以便对基于场景的音频节目的响度进行测量。

注1 – 用户应该知道所测得的响度是对于主观响度的一个估计，取决于收听者、音频素材和收听条件，所测得的响度含有一定程度的不确定性。

注2 – 对于符合本建议书的仪表测试合规性，可使用ITU-R BS.2217号报告所述数据集中的测试材料。

附件1

客观多声道响度测量算法的详细说明

本附件规定了多声道响度测量建模算法。

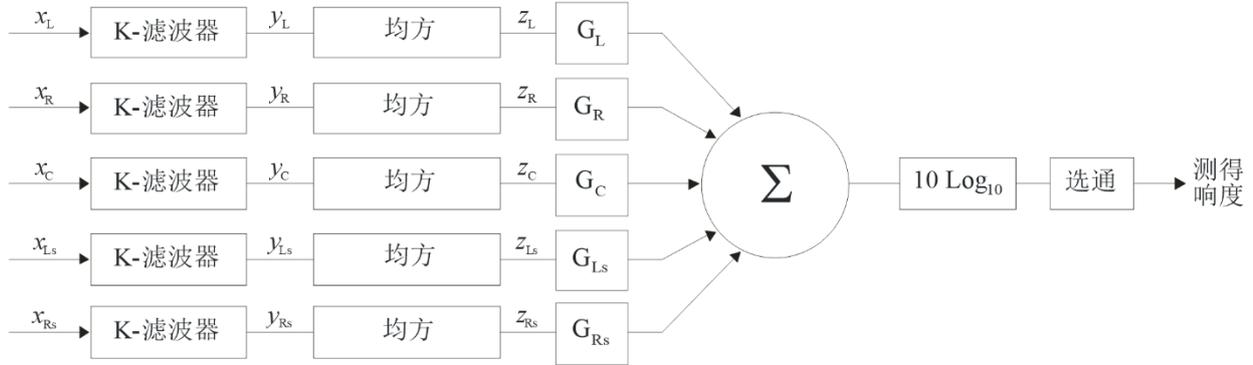
该算法由四级构成：

- “K” 频率加权；
- 每声道均方计算；
- 信道加权和（环绕信道加权值更大，并不包括LFE信道）；
- 400 ms块选通（75%叠加），其中采用了两个门限：
 - 第一个是 -70 LKFS；
 - 第二个是应用于第一个门限后相对于测得电平的 -10 dB。

图1所示的是该算法各个组成部件的方框图。为了帮助描述该算法，沿着信号流程在不同点作了标记。方框图显示了五个主要声道的输入（左边、中间、右边、左边环绕和右边环绕）；这样允许监听包含一至五个声道的节目。对于少于五个声道的节目，一些输入将不会使用。本测量法中不包括低频效应（LFE）声道。

图1

多声道响度算法简略方框图

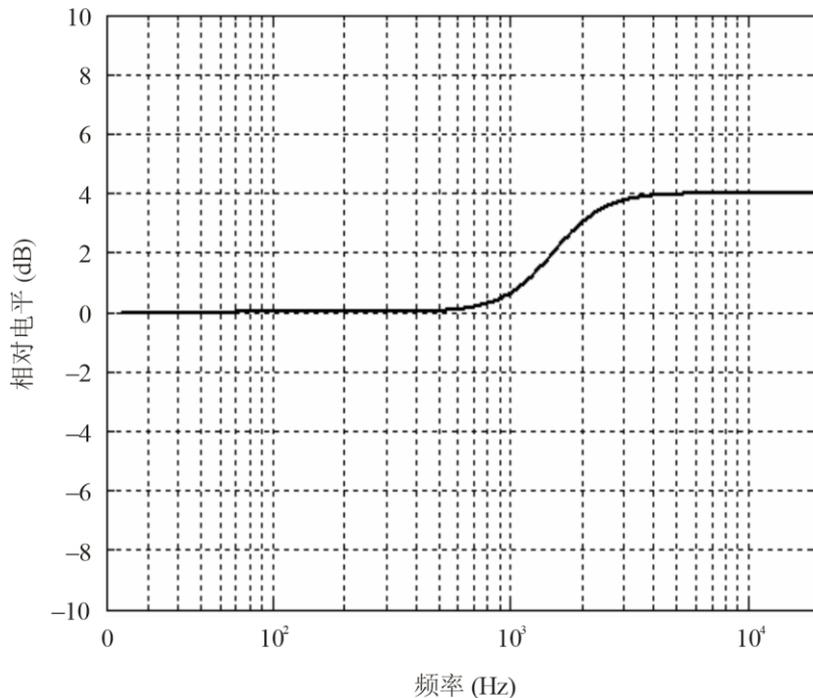


BS.1770-01

算法的第一步采用信号的2级前置滤波器¹。前置滤波器第1级计入了前端响度效应（前端被模拟为刚性球体）。响应如图2所示。

图2

用于说明前端声学效应的前置滤波器第1级响应



BS.1770-02

¹ K加权滤波器由滤波器的两级构成：第1级的搁置滤波和第2级的交通滤波。

前置滤波器第1级定义为如图3所示的滤波器，滤波器系数见表1。

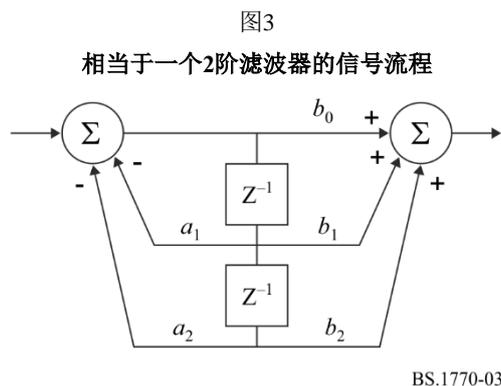


表1

模拟球体前端的前置滤波器第1级的滤波器系数

		b_0	1.53512485958697
a_1	-1.69065929318241	b_1	-2.69169618940638
a_2	0.73248077421585	b_2	1.19839281085285

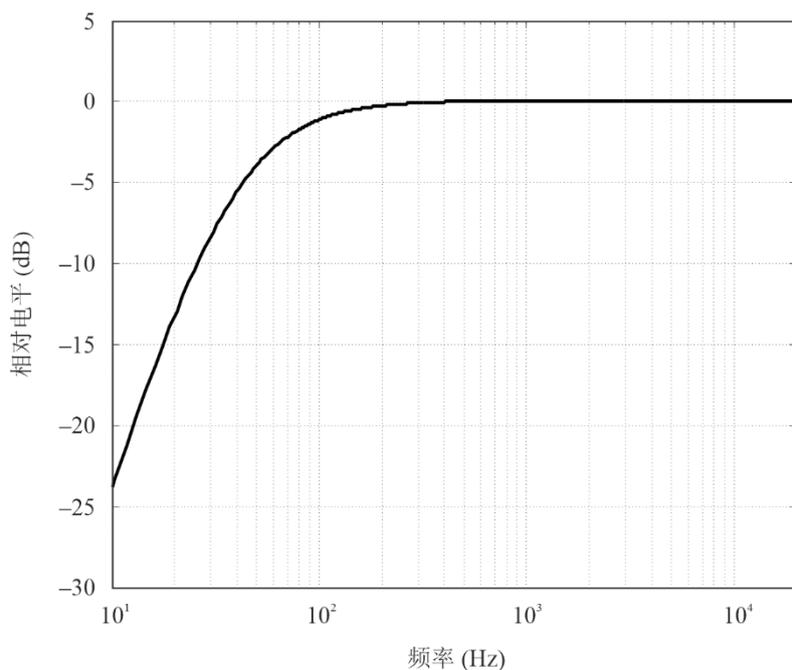
这些滤波器系数是针对48 kHz的采样率，采用其它采样率将需要不同的系数值，这些系数值应经过选择以便提供的频率响应与指定的滤波器在48 kHz采样率提供的频率响应相同。由于可获得硬件的固有精度，这些系数值可能需要进行量化。测试已经表明该算法的性能对于这些系数的微小变化不敏感。

前置滤波器的第二级应用如图4所示的一个简单高通滤波器。

该级的加权曲线定义为如图3所示的一个2阶滤波器，其系数见表2。

这些滤波器系数是针对48 kHz的采样率，采用其它采样速率将需要不同的系数值，这些滤波器系数应该经过选择以便提供的频率响应与指定的滤波器在48 kHz采样率提供的频率响应相同。

图4
第二级加权曲线



BS.1770-04

表2
第二级加权曲线的滤波器系数

		b_0	1.0
a_1	-1.99004745483398	b_1	-2.0
a_2	0.99007225036621	b_2	1.0

测量间隔 T 内过滤输入信号的幂和均方计量如下：

$$z_i = \frac{1}{T} \int_0^T y_i^2 dt \quad (1)$$

其中， y_i 是输入信号（由上述两级前置滤波器过滤，以及 I ，其中 $I = \{L, R, C, Ls, Rs\}$ ，）一套输入声道。

整个测量间隔 T 内的响度定义为：

$$\text{响度, LK} = -0.691 + 10 \log_{10} \sum_i G_i \cdot z_i \quad \text{LKFS} \quad (2)$$

其中 G_i 是单个声道的加权系数。

为计算选通响度测量值，间隔 T 被分为一套叠加的选通块间隔。一个选通块是持续 $T_g = 400$ ms的连续音频采样（至最近的采样）。每一选通块的叠加须为选通块持续时间的75%。

须限制测量间隔，以使其在选通块结尾处结束。测量间隔结束时的不完整选通块不被使用。

间隔 T 中第 i 个输入声道的第 j 个选通块的幂和均方为：

$$z_{ij} = \frac{1}{T_g} \int_{T_g \cdot j \cdot \text{step}}^{T_g \cdot (j \cdot \text{step} + 1)} y_i^2 dt \quad \text{其中步长} = 1\text{-叠加}$$

$$\text{及 } j \in \left\{ 0, 1, 2, \dots, \frac{T - T_g}{T_g \cdot \text{step}} \right\} \quad (3)$$

第 j 个选通块响度定义为：

$$l_j = -0.691 + 10 \log_{10} \sum_i G_i \cdot z_{ij} \quad (4)$$

对选通门限 Γ ，有一套选通块指标 $J_g = \{j : l_j > \Gamma\}$ ，其中选通块响度高于选通门限。 J_g 中的元素数为 $|J_g|$ 。

因此，测量间隔 T 的被选通响度定义为：

$$\text{被选通响度 } L_{KG} = -0.691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) \text{LKFS} \quad (5)$$

用两级程序进行被选通测量：首先用绝对门限，然后用相对门限。低于绝对门限的选通块不用于相对选通的计算。通过使用绝对门限 $\Gamma_a = -70$ LKFS测量响度的方法计算相对门限 Γ_r ，然后再从中减去10，因此，

$$\Gamma_r = -0.691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) - 10 \text{ LKFS}$$

其中：

$$J_g = \{j : l_j > \Gamma_a\}$$

$$\Gamma_a = -70 \text{ LKFS} \quad (6)$$

然后采用 Γ_r 计算被选通响度：

$$\text{被选通响度 } L_{KG} = -0.691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) \text{LKFS}$$

其中：

$$J_g = \{ j : l_j > \Gamma_r \text{ and } l_j > \Gamma_a \} \quad (7)$$

在该测量中使用的频率加权，由前置滤波器生成（1级滤波级联）及RLB加权，并定名为“K”加权。等式（2）中计算的响度值的数字结果之后应有“LKFS”标志，表示“K加权的相对满刻度的响度”。该LKFS单位等于一个分贝，其中信号电平每增加1 dB，将导致响度数值增加1LKFS。

如果一个0 dB FS的1 kHz（确切为997Hz，见注1和2）正弦波从左侧、中间或右侧应用于声道输入端，所指的响度将等于-3.01 LKFS。

注1 – 等式（2）中的恒量-0.691抵消了997Hz的K加权增益。

注2 – IEC61606指出，除非另有规定，测量参考频率须为实际频率997Hz。在非关键情况下，该频率可能标称为1KHz。

表3给出了每一信道的加权系数。

表3
各个音频信道的加权

通道	加权, G_i
左边(G_L)	1.0 (0 dB)
右边(G_R)	1.0 (0 dB)
中间(G_C)	1.0 (0 dB)
左边环绕(G_{Ls})	1.41 (~ +1.5 dB)
右边环绕(G_{Rs})	1.41 (~ +1.5 dB)

应注意，虽已经证实这个算法能有效地用于音频节目（这是典型的广播内容），但这个算法通常不适用于估算纯单音的主观响度。

附件1 的后附资料1 (资料性)

多声道测量算法的描述及发展

本后附资料描述了一个新近开发的、用于客观地测量音频信号的感觉响度的算法。该算法能够用来准确地测量单声道、立体声和多声道信号的响度。被提议算法的关键好处是它的简易性，允许以极低的代价来实现它。本后附资料也描述了为了形成一个主观数据库而进行正式主观测试的结果，该数据库用于评估该算法的性能。

1 引言

有许多应用需要测量和控制音频信号的感觉响度。这种应用的例子包括电视和无线电广播应用，在这种情况下音频素材的性质和内容经常发生变化。在这些应用中，音频的内容能够在音乐、语音和音响效果或者这些声音的某些组合之间频繁地转换。在节目素材内容方面的这种变化能够导致主观响度的显著改变。此外，各种各样形式的动态处理常常应用于这些信号，这能够对信号的感觉响度产生显著的影响。当然，主观响度对于音乐产业也极为重要，在这里，通常采用动态处理使一段录音的感觉响度最大化。

近年来，在无线电通信第6P工作组内部已经有一项正在进行的工作，为广播应用确定一种测量典型节目素材的感觉响度的客观方法。ITU-R第一阶段的工作是专门调查客观的单声道响度算法，已经证实加权的均方测量 $Leq(RLB)$ 能够为单声道信号提供最佳的性能[A1-2]。

充分认识到的是，对于广播应用，需要能够用于单声道、立体声和多声道信号的响度测量。本后附资料提出了一种新的、能够成功地用于单声道、立体声和多声道音频信号的响度测量算法。被提议的算法以 $Leq(RLB)$ 算法的直接扩展作为基础，此外，新的多声道算法保留了单声道 $Leq(RLB)$ 算法极低的计算复杂度。

2 背景

在ITU-R研究的第一阶段，开发出一种主观测试方法来检查典型单声道节目素材的响度感觉[A1-2]。为了建立一个用于评估可能的响度测量算法性能的主观数据库，要在全世界的五个地点进行主观测试。被测试者比较各种单声道音频片段和一个参考片段的响度，音频序列取自实际的广播素材（电视和收音机）。

结合这些测试，在加拿大通信研究中心的音频感觉实验室，七个不同的建议者提交了共计十个商业上开发的、用于评估的单声道响度计/算法。

另外，Soulodre提交了两个附加的基本响度算法作为性能基线[A1-2]。这两种客观测量包括一个简单的频率加权函数，接着是一个均方测量方框。这两种测量方法之一的 $Leq(RLB)$ 采用的是被称之为修正低频B曲线(RLB)的高通频率加权曲线。

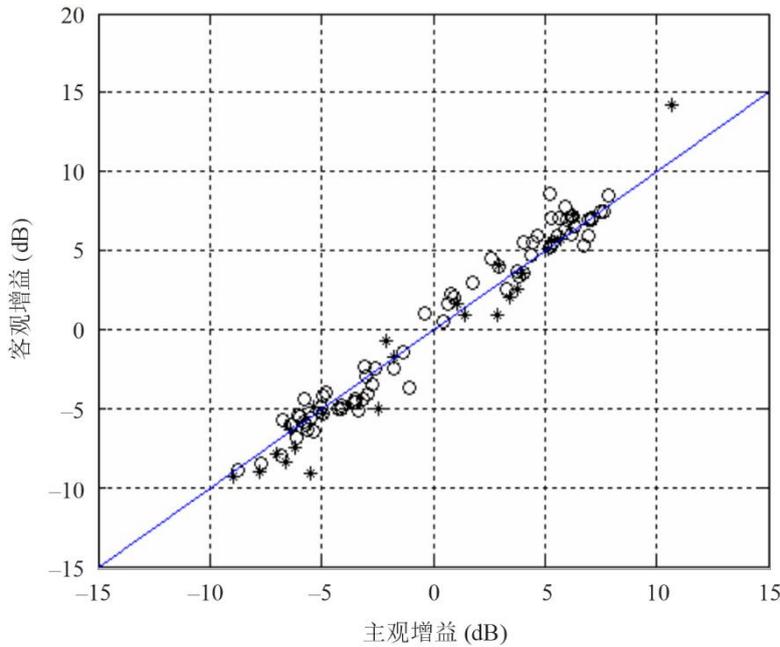
另一种测量方法 Leq 是简单的没有加权的均方测量。

图5所示的是关于 $Leq(RLB)$ 响度计的最初ITU-R研究结果。横轴表示的是来自主观数据库的相对主观响度，而纵轴表示的是由 $Leq(RLB)$ 测量预测的响度。图中每一个点代表测试中一个音频测试片段的的结果。圆圈代表基于语音的音频片段，而星号代表不是基于语音的片段。可以看出数据点紧紧地聚集在对角线周围，显示了 $Leq(RLB)$ 响度计有很好的性能。

在所有被评估的响度计中，发现 $Leq(RLB)$ 能够提供最好的性能（虽然在统计显著性领域里，最好还是实行某些基于心理声学的计量），发现 Leq 表现的性能几乎和RLB一样好。这些发现表明对于典型的单声道广播素材，相对于更加复杂的、可能包含详细感觉模型的度量，简单的基于能量的测量同样地强健。

图5

单声道Leq(RLB)响度计对比主观结果 (r = 0.982)



BS.1770-05

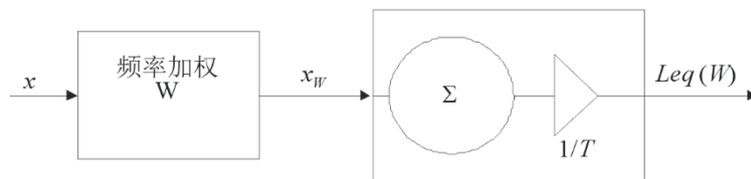
3 Leq(RLB)算法设计

特地将Leq(RLB)响度算法设计得非常简单，Leq(RLB)算法的方框图如图6所示，它包括一个高通滤波器，接着是一个在时间上求能量均值的方框。滤波器的输出转到一个求能量的总和而且计算对于时间的均值的处理方框。

这个滤波器的作用是提供一些信号频谱含量的感觉相关加权。对于响度测量采用这种基本结构的一个好处是，以具有极低计算要求的简单时域单元就能够完成所有的处理。

图6

简单的基于能量响度测量方框图



BS.1770-06

图6所示的Leq(RLB)算法只是等效声级 (Leq) 测量的一个频率加权形式，Leq定义如下：

$$Leq(W) = 10 \log_{10} \left[\frac{1}{T} \int_0^T \frac{x_W^2}{x_{Ref}^2} dt \right] \quad \text{dB} \quad (8)$$

其中：

x_w : 在加权滤波器输出端的信号

x_{Ref} : 一些参考电平

T : 音频片段的长度。

$Leq(W)$ 中的符号 W 代表频率加权，在这种情况下它是修正的低频B曲线(RLB)。

4 主观测试

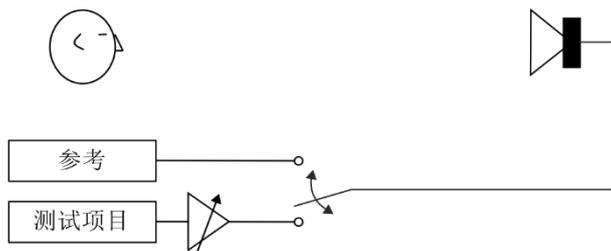
为了评估可能的多声道响度测量，必须进行正式的主观测试从而建立一个主观数据库。于是，能够以它们预测主观测试结果的能力来评估可能的响度测量算法。这个数据库提供了各种各样单声道、立体声和多声道节目素材的感觉响度评定。在测试中采用的节目素材取自全世界真实的电视和无线电广播，以及CD和DVD。片段包括音乐、电视剧和电影、运动事件、新闻广播、音响效果和广告，片段中包含的是多种语言的语音片段。

4.1 主观测试设置

主观测试由响度匹配任务构成。被测试者收听范围广泛的典型节目素材，调整各个测试项目的电平直到它的感觉响度与参考信号的感觉响度相当（见图7）。

参考信号总是以60 dBA电平重复产生，这个电平由Benjamin所确定，是在实际家中收看电视的典型收听电平 [A1-1]。

图7
主观测试方法

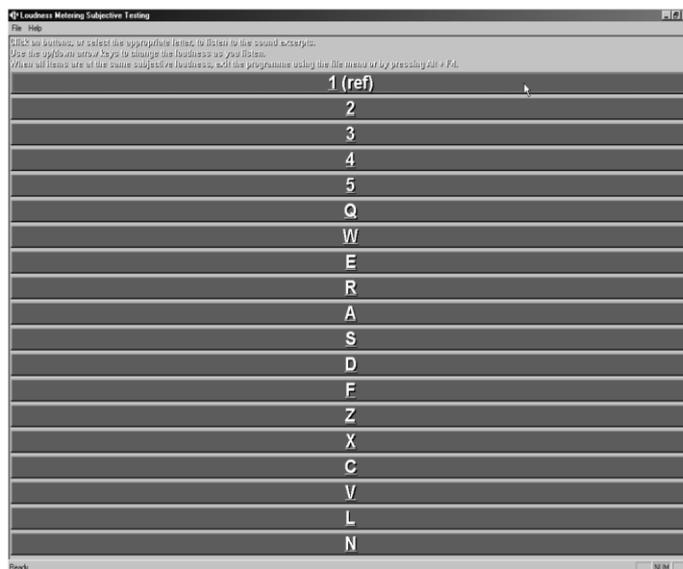


BS.1770-07

基于软件的多声道主观测试系统，由澳大利亚广播公司开发并提交，允许收听者在测试项目之间即时地来回切换，并调整各个项目的电平（响度）。测试软件的屏幕截图如图8所示，测试项目的电平能够以0.25 dB的步长进行调整。选择标记为“1”的按钮接入参考信号，参考信号的电平保持固定不变。

图8

主观测试系统的用户界面



BS.1770-08

被测试者使用计算机键盘选择一个特定的测试项目，调整它的电平直到它的响度与参考信号相当。被测试者通过选择合适的关键字，能够在任意的测试项目之间即时地切换。片段在测试期间连续地（循环地）播放，软件记录下被测试者为各个测试项目调整的增益设置。因此，主观测试产生一组使各个测试片段的响度与参考片段相当所需要的增益值（分贝数），这样允许直接地确定各个测试项目的相对响度。

在进行正式的不知情的测试之前，各个被测试者要经过一段时间的训练，在此期间他们要熟悉测试软件以及他们在实验中的任务。由于许多测试项目包含了语音和其它声音的混合体（即音乐、背景噪声，等等），要特别地指导被测试者去匹配整个信号的响度，而不仅仅是信号中的语音部分。

在正式的盲试中，提供给各个被测试者的测试项目的顺序是随机的，因而，没有两个被测试者会被提供顺序相同的测试项目，这样做是为了消除由于顺序效应引起的任何可能的偏差。

4.2 主观数据库

用于评估被提议算法性能的主观数据库实际上由三个单独的数据集组成，数据集的建立来自三个独立的、进行了数年的主观测试。

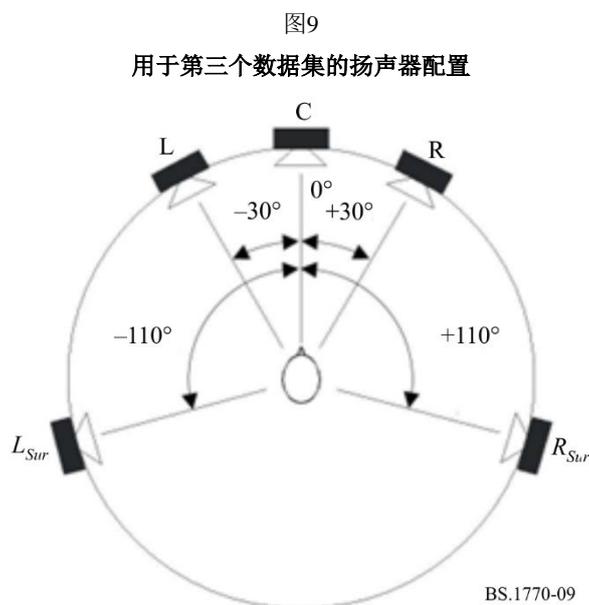
第一个数据集由最初ITU-R研究的结果组成，在这个研究中，被测试者要匹配96个单声道音频片段的感知响度。对于这个数据集，主观测试是在全世界五个单独的地点提供了共计97名收听者而进行的。由无线电通信WP 6P SRG3成员组成的一个三人专门小组挑选测试片段以及参考项目。本实验中的参考信号由英语女性语音组成。片段由直接放置在收听者前面的一个扬声器播放。

在最初的ITU-R单声道研究之后，一些算法的建议者推测用于主观测试的信号的范围和类型不够广泛，他们进一步推测由于这个原因，简单的 Leq (RLB)基于能量的算法要优于所有的其它算法。

为了解除这种担忧，要求建议者提交用于进一步主观测试的新的音频片段，鼓励他们提交他们觉得对于 $Leq(RLB)$ 算法更具挑战性的单声道片段，只有两个响度计的建议者提交了新的片段。

在加拿大通信研究中心的音频感觉实验室，使用这些新的序列进行了正式的主观测试。共计20名被测试者提供了96个单声道片段的响度评定，测试采用用于建立第一个数据集的相同的主观方法，也使用相同的参考信号。这些测试的结果形成了主观数据库的第二个数据集。

第三个数据集包括144个音频片段的响度评定，测试序列由48个单声道项目、48个立体声项目和48个多声道项目组成。此外，有一半的单声道项目由中间的通道（单声道）播放，而另一半的单声道项目由左边和右边的扬声器（两个单声道）播放。这样做是为了说明一个人听单声道信号可能采取的两种不同的方式。对于这个测试，参考信号由具有立体声环境和低电平背景音乐的英语女性语音组成。共计20名被测试者参与了这次测试，该测试采用ITU-R BS.775建议书规定的扬声器配置，如图9所示。



前两个数据集限于单声道测试片段，因此，镜像不是一个因素。在第三个数据集中，还包含了立体声和多声道序列，镜像是一项需要处理的重要考虑事项。有人认为在一个片段内的镜像和环境很有可能对于片段的感受响度有着显著的影响。因此，为了包含范围广泛的镜像类型（例如，中间对比，前置声源对比环绕声源）和变化的环境总额（例如，不加渲染的对比回响的），要选择立体声和多声道片段。

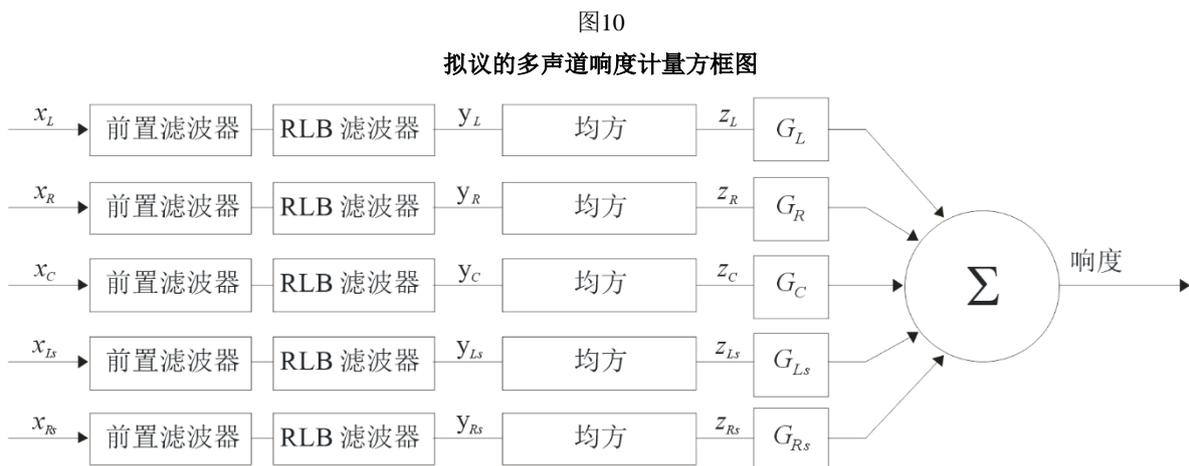
被测试者必须同时匹配单声道、双声道、立体声和多声道信号的情况意味着，这个测试在本质上比以前的限于单声道信号的数据集更加困难，各种镜像类型和变化的环境总额加剧了这种困难。曾有一些顾虑是，由于这些因素，被测试者可能会被这项任务所难倒。幸运的是，初步的测试表明这项任务是可以完成的，20名被测试者确实能够提供一致的结果。

5 多声道响度算法设计

如前所述，*Leq*(RLB)算法被设计成对于单声道信号有效，早期的研究已经表明，对于这项任务该算法是十分成功的。多声道响度算法的设计带来了一些额外的挑战，对一个成功的多声道算法的关键要求是它必须也能很好地用于单声道、双声道和立体声信号。就是说，这些形式必须被当作多声道信号的一个特例（即便是非常普通的情况）。

在当前的研究中，假设多声道信号符合标准ITU-R BS.775建议书5.1通道配置，对于LFE通道不予说明。

在多声道响度计中，每个单独的音频通道的响度是独立地采用*Leq*(RLB)单声道算法测量的，如图10所示。然而，在*Leq*(RLB)测量之前要对各个通道进行前置滤波。



BS.1770-10

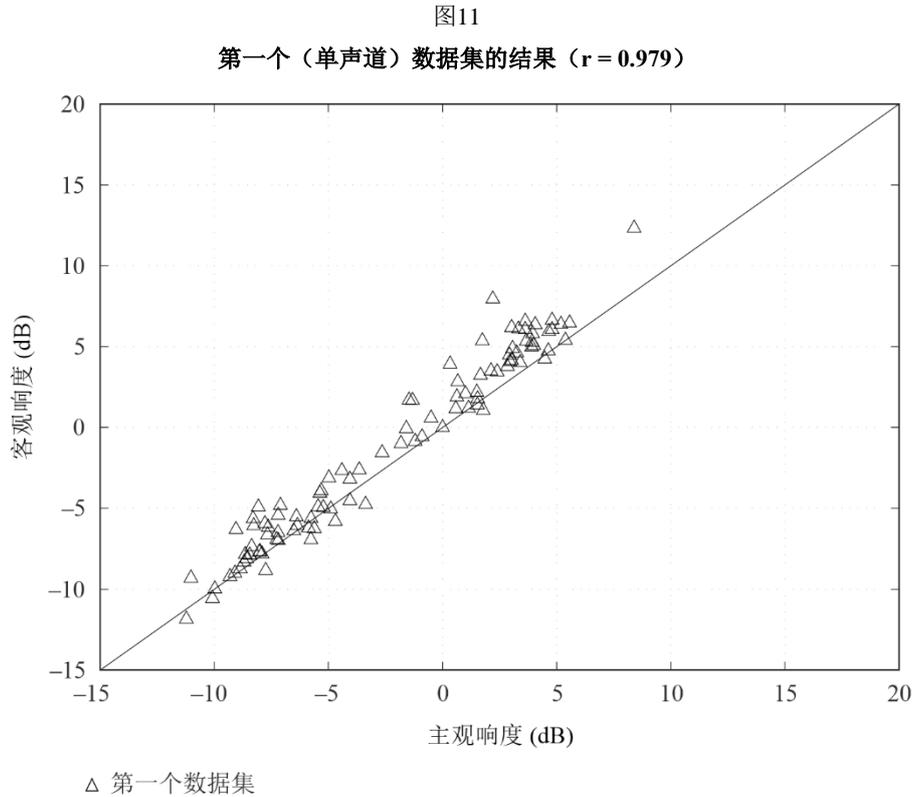
前置滤波器的作用是为了说明有输入信号的头部的声学效应，这里，将头部模拟成一个固定球体。各个通道采用相同的前置滤波器，根据信号到达的角度对最终的响度值进行加权（ G_i ），然后求和（在线性域）从而提供一个合成的响度测量。加权用于考虑到从收听者后面到达的声音可能感觉要比从收听者前面到达的声音响亮的情况。图10中的“前置滤波器”和“RLB滤波器”组合指的是上文附件1主要部分说明的K加权。

拟议的多声道响度算法的关键好处是它的简单性，该算法完全由非常基本的信号处理单元组成，这些单元能够在时域用廉价的硬件轻松地实现，该算法的另一个关键好处是它的可量测性，由于各个通道采用的处理相同，实现一种能够包含1到N任意数量通道的响度计是非常直接的。此外，由于以各个通道作用的总和作为响度值，而不是信号电平，算法不依赖于通道之间的相位或相关性。这使得被提议的响度量更加通用和强健。

6 多声道算法评估

三个数据集中采用的336个音频片段要经过被提议的多声道算法的处理，并记录下预测的响度评定。由于这一步处理，该算法的总性能能够以预测的评定和真实的主观评定之间的一致性作为基础来进行评估。

图11、12和13描绘的是被提议的、适用于三个数据集的响度计的性能。在每幅图中，横轴给出的是数据集中每个音频片段的主观响度，纵轴表示的是由被提议的响度计所预测的客观响度。图中各点代表各个音频片段的的结果。应该注意到一个最佳的客观算法将让所有的数据点都落在斜率为1且经过原点的对角线上（如图所示）。

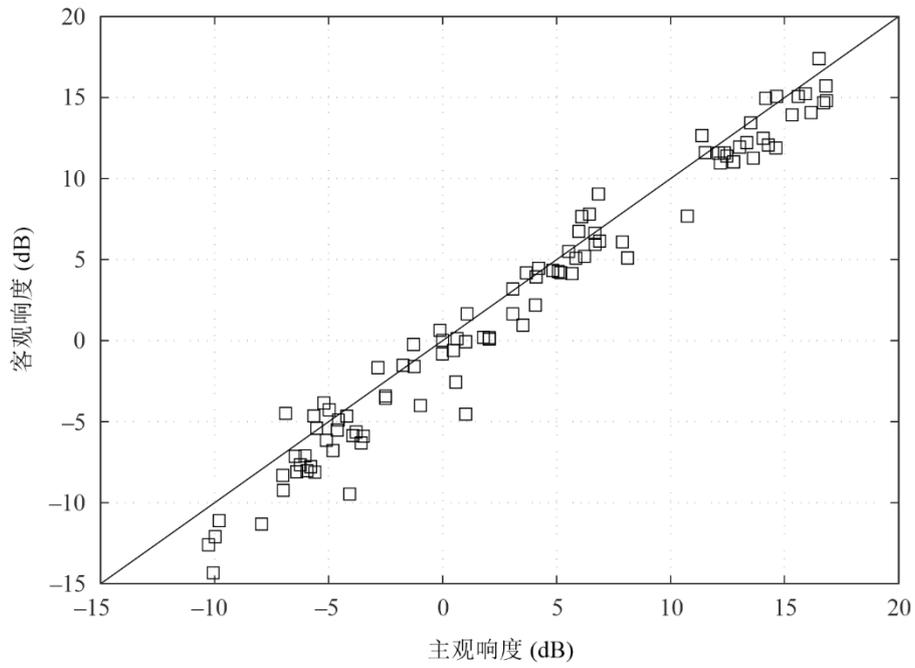


从图11中可以看出，被提议的多声道响度算法在预测第一个（单声道）数据集的结果方面表现非常良好，主观响度评定与客观响度量度之间的相关性为 $r = 0.979$ 。

如图12所示，对于第二个数据集，主观响度评定与客观响度量度之间的相关性也非常得好（ $r = 0.985$ ）。有趣的是注意到在这个数据集中差不多一半的片段为音乐。

图12

第二个（单声道）数据集的结果 ($r = 0.985$)

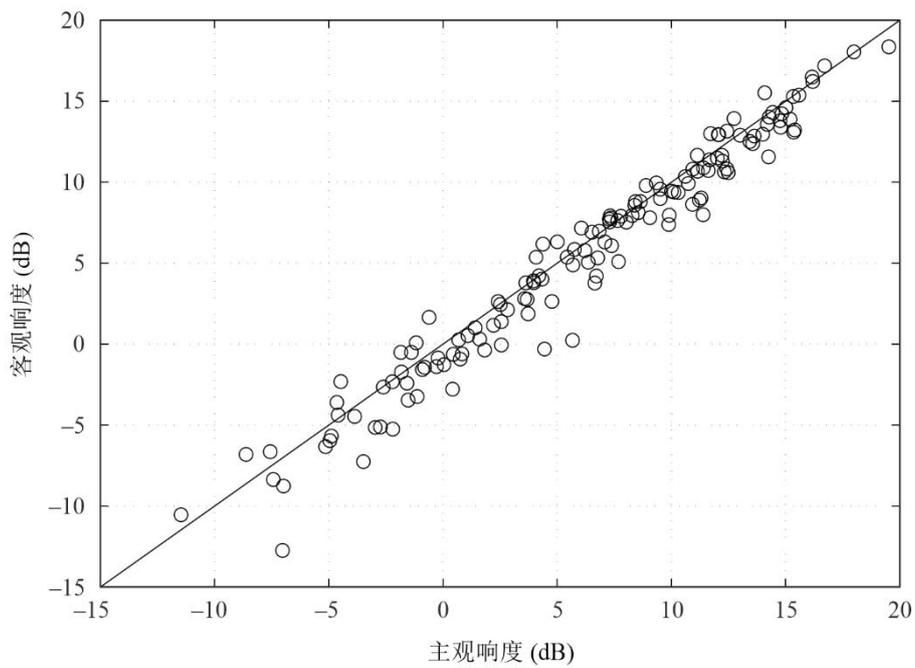


□ 第二个数据集

BS.1770-12

图13

第三个（单声道、立体声和多声道）数据集的结果 ($r = 0.980$)

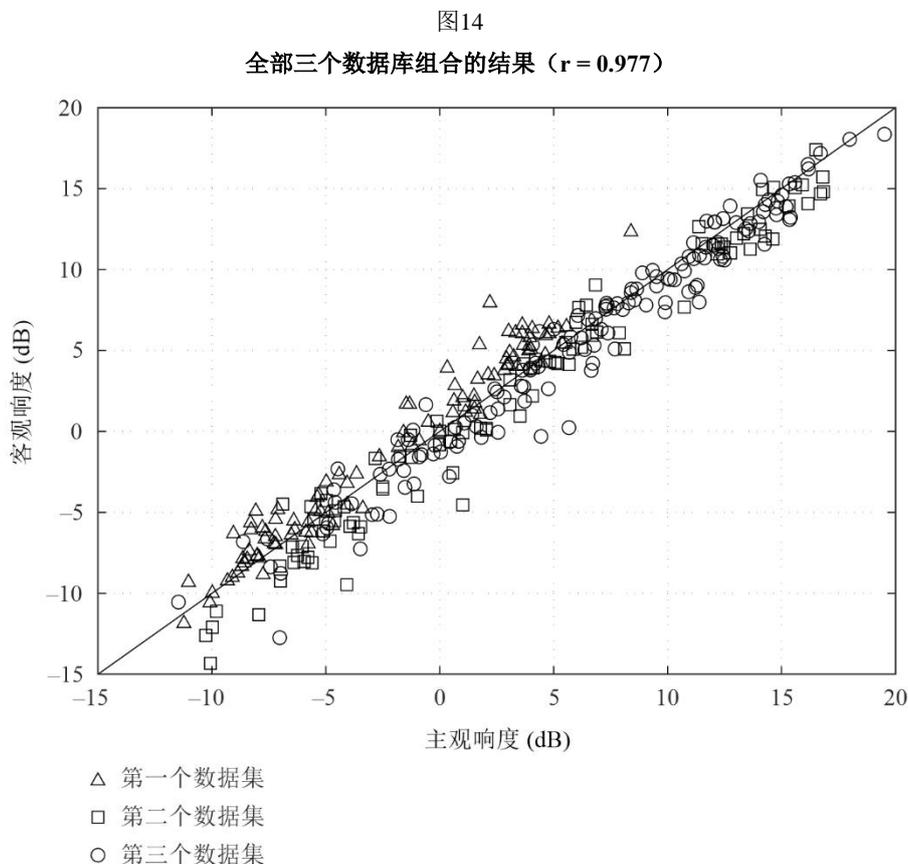


○ 第三个数据集

BS.1770-13

图13所示是第三个数据集的结果，该数据集包括单声道、双声道、立体声和多声道信号。包含在图13和图14中的多声道结果是针对指定的算法，但是将环绕声道加权设为4 dB（最初的提议）而不是1.5dB（最后的规范）。已经证实4.0 dB到1.5 dB的变化对于结果不会产生任何显著的影响。同样地，该算法的性能非常好，相关性 $r = 0.980$ 。

检查算法的性能对于组成主观数据库的所有336个音频序列是有益的。因此，图14组合了来自三个数据集的结果。可以看出对于整个主观数据库性能都是非常得好，总的相关性 $r = 0.977$ 。



BS.1770-14

这次评估的结果表明基于 $Leq(RLB)$ 响度测量的多声道响度量算法，对于主观数据库的全部336个序列均表现非常好。主观数据库提供了范围广泛的节目素材，包括音乐、电视剧和电影、运动事件、新闻广播、音响效果和广告。片段中也包含了多种语言的语音，此外，该结果表明了被提议的响度计适用于单声道、立体声以及多声道信号。

参考文献

- [A1-1] BENJAMIN, E. (2004年10月) 家庭环境中对话再现的首选聆听水平和接受范围，音频工程学会第117届大会，旧金山，预印本6233。
- [A1-2] SOULODRE, G.A. (2004年5月) 客观响度表的评估，音频工程学会第116届大会，柏林，预印本6161。

附件2

准确测量“真正峰值”电平的导则

本附件描述了在一个单通道线性PCM数字音频信号范围内评估真正峰值电平的算法。接下来的讨论假定采样率为48 kHz，真正峰值电平是信号波形在连续时域中的最大值（正的或负的）；这个数值可能要大于48 kHz采样时间范围内的最大样本。

1 摘要

处理的步骤为：

- 1 衰减：12.04 dB的衰减
- 2 4 × 过采样
- 3 低通滤波器
- 4 绝对的：绝对值
- 5 转变为dB TP

2 方框图



3 详细说明

第一步包括实施12.04 dB的衰减（2 bit一档）。这一步的目的是为后续信号处理采用整数运算提供空间。如果采用浮点进行计算，则不需要这一步。

4 × 过采样滤波器把信号采样率从48 kHz增加到192 kHz，信号的这种高采样率形式能够更加准确地表示音频样本内部所代表的真实波形，较高的采样率和过采样比更为可取（见本附件的后附资料1），较高采样率的输入信号要求按比例降低过采样（例如，对于采样速率为96 kHz的输入信号，2 × 过采样就足够了）。

满足要求的一组滤波器系数（顺序为48 kHz、4相、FIR内插）如下所示：

相位0	相位1	相位2	相位3
0.0017089843750	-0.0291748046875	-0.0189208984375	-0.0083007812500
0.0109863281250	0.0292968750000	0.0330810546875	0.0148925781250
-0.0196533203125	-0.0517578125000	-0.0582275390625	-0.0266113281250
0.0332031250000	0.0891113281250	0.1015625000000	0.0476074218750
-0.0594482421875	-0.1665039062500	-0.2003173828125	-0.1022949218750
0.1373291015625	0.4650878906250	0.7797851562500	0.9721679687500
0.9721679687500	0.7797851562500	0.4650878906250	0.1373291015625
-0.1022949218750	-0.2003173828125	-0.1665039062500	-0.0594482421875

0.0476074218750	0.1015625000000	0.0891113281250	0.0332031250000
-0.0266113281250	-0.0582275390625	-0.0517578125000	-0.0196533203125
0.0148925781250	0.0330810546875	0.0292968750000	0.0109863281250
-0.0083007812500	-0.0189208984375	-0.0291748046875	0.0017089843750

样本的绝对值通过对负数的样本取相反数得到；在这一点上信号是单极性的，负数用相同幅度的正数代替。

四个阶段（衰减、过采样、过滤和取绝对值）后的结果是一个数字，与原始样本值（例如，24位整数）在相同的域。在这之后，有必要把初始的12.04 dB衰减补上。这使得处理过程的总增益正常化至统一。

必须要知道的是，通常，将衰减值增大12分贝（2位左移）要求将衰减值转换到一个数字格式，该数字格式能够代表比原始格式的标称满度范围更高的值。可以选择适当校准表头刻度的方式来扩大结果。

遵循这些指导原则，且采用至少192 kHz过取样的采样率的计算表，应以“dB TP”为单位显示结果，已将结果转换成一个对数刻度。这可以通过计算衰减值、过采样值、过滤值、绝对值的 $20\log_{10}$ ，再加上12.04 dB来实现。“dB TP” .Designation名称表示相对于100%满刻度和真峰值测量的分贝。

附件2 的后附资料¹² (资料性)

关于数字音频信号的准确峰值计量的考虑

问题是什么？

数字音频系统中的峰值计通常记录下“峰值样本”，而不是“真正峰值”。

峰值样本计通常是各个输入样本的绝对（经过整流的）值与峰值计的当前读数进行比较；如果新样本较大，则它取代当前的读数；否则，当前的读数乘以一个略小于1的常数，产生一个对数衰减。由于易于实现，这种峰值计普遍存在，但是它们不一定能够记录下音频信号的真正峰值。

因而，在准确计量节目峰值较为重要的地方采用峰值样本计可能会带来问题。不幸的是，绝大多数的数字峰值计是峰值样本计，尽管对于操作者而言这通常不明显。

问题之所以出现是由于采样信号的真正峰值通常出现在样本之间而不是精确地位于采样时刻，像这样的峰值不会被峰值样本计正确地记录下来。

² 注1 – 下列资料性文本由AES标准工作组SC-02-01通过无线电通信关于响度计量的WP 6J报告人提交。

这会导致许多类似的峰值样本计异常：

- 不一致的峰值读数：经常注意到为采用峰值样本计的数字系统重复地播放一段模拟录音，每次播放都会产生完全不同的节目峰值读数。同样地，如果在计量之前，经过一个采样速率转换器重复地播放一段数字录音，每次播放记录下的峰值也不同。这是因为在每一次播放的时候，采样时刻可能落在实际信号的不同部分。
- 突然的过载：由于采样信号可能包含过载，甚至当它们在到达或者甚至接近数字满标而没有样本的时候，峰值计的过载指示是不可靠的，过载可能导致后续过程中的限幅，例如在特殊的D/A变换器内部或者采样率转换期间，即使先前它们没有被峰值样本计记录下来（在那个点进行监测时甚至听不见它们）。
- 被计量音调的读数不足和差拍：接近于采样频率整数因子的单音（例如校准音）可能读数不足，或者即使音调的幅度固定不变可能产生一个不断变化的读数。

问题能有多严重？

通常，峰值样本计量的信号的频率越高，可能的误差就越大。

对于连续的单音，较容易证实，例如，对于频率为四分之一采样频率、相位不适宜的音调，读数不足为3 dB。对于频率为二分之一采样频率的音调读数不足可能几乎是没有限定的；然而，绝大多数的音频信号在这个频率点不包含太多的能量（由于它的大部分都被D/A变换器的抗混叠滤波器滤除了，而且由于“实际的”声音通常不会由连续的高频占主要地位）。

由于差频（ nf_{tone} 和 f_s 之间的差）相对于计量衰减率的倒数较高，频率不靠近采样频率的最小整数因子的连续音调在峰值样本计中不会读数不足。换言之，采样时刻充分地接近于音调的真正峰值时，峰值计常常不会读数不足。

然而，对于个别过渡状态，上面的那个机制不能掩盖读数不足，于是过渡状态的频率含量越高，可能的读数不足就越大。在“实际的”声音中出现具有显著高频含量的过渡状态是正常的，这些高频含量的读数不足通常为几分贝。

由于实际声音的频谱通常随着频率的增长而降低，而且由于这一点不会随着采样频率的增加而变化，采用较高的原始采样频率时，峰值样本计读数不足会锐减。

解决方案是什么？

为了计量被采样信号的真实峰值，有必要对信号进行“过采样”（或者“超采样”）本质上恢复已有样本之间的原始信号，因而提高信号的采样率。这个建议听起来可疑：如何能够恢复那些看来似乎已经丢失的信息？事实上，采样理论表明这一点能够做到，因为已知被采样的信号所包含的频率不超过原始采样频率的一半。

多大的过采样比是必需的？要找到答案需回答几个问题：

- 最大的、可接受的读数不足误差是多少？
- 被计量的最高频率与采样频率（最大的“归一化频率”）的比是多少？

如果已知这些准则，就有可能通过一种直接的“方格纸”方法来计算所需要（甚至不用考虑过采样实现的细节）的过采样比。可以简单地认为，在我们的最大归一化频率点，对称地出现在正弦曲线波峰两侧的一对过采样样本将导致读数不足，这是“最坏情况”的读数不足。

因而对于：过采样比， n

最大的归一化频率， f_{norm}

采样频率， f_s

可以看到：

按照过采样速率的采样周期为 $1/n \cdot f_s$

最大归一化频率的周期为 $1/f_{norm} \cdot f_s$

则：

最大的读数不足（dB）为 $20 \cdot \log(\cos(2 \cdot \pi \cdot f_{norm} \cdot f_s / n \cdot f_s \cdot 2))$

（分母中的2是因为可能最多错过一个波峰达半个采样周期）

或者：

最大的读数不足（以分贝计）= $20 \cdot \log(\cos(\pi \cdot f_{norm} / n))$

这个公式用于构造下表，该表很可能涵盖了被关注的范围：

过采样比	读数不足（dB）最大值 $f_{norm} = 0.45$	读数不足（dB）最大值 $f_{norm} = 0.5$
4	0.554	0.688
8	0.136	0.169
10	0.087	0.108
12	0.060	0.075
14	0.044	0.055
16	0.034	0.042
32	0.008	0.010

应怎样实现一个真正峰值计？

为了以要求的过采样速率产生数据流，过采样操作是在原始的样本之间插入数值为0的样本，然后应用低通“内插”滤波器滤除所要求的最大 f_{norm} 以上的频率。如果现在对过采样信号运用峰值样本算法，会得到具有所要求的最大读数不足的真正峰值计。

考虑这样一个过采样器的实现是有趣的。这种低通滤波通常被当作一个对称的FIR来实现。在采用这种滤波器传送高质量音频的地方，例如在过采样D/A变换器或者采样率转换器中，为了保持极低的通带纹波、获得最大的阻带衰减和狭窄的过渡带，需要计算大量的“抽头”。为了保持动态范围且使失真最小化，也必须保留较长的字长。

然而，由于不会去收听我们的过采样器的输出，而只用它显示一个读数或者产生一个柱状图，可能不会得到同样的精确度要求。只要通带纹波，加上附加的来自阻带的寄生分量，不会降低读数精度以至于超出我们的目标，结果就会令人满意，这会大大地减少必需的抽头数量，尽管取决于我们的最大归一化频率指标，可能仍然需要获得一个狭窄的过渡带。同样地，为了保证我们的目标精度降至柱状图的底部，可能只需要足够的字长，除非对于较低的幅度要求准确的数字输出。

于是，有可能适当的过采样器（可能适用于许多通道）可以用普通的廉价DSP或者FPGA或者可能更普通的处理器实现。另一方面，采用为D/A变换器应用设计的高精度过采样芯片已经实现了过采样峰值计，虽然这相当浪费硅片和功率，但是器件是廉价的和易于得到的。

对于一种特定的峰值计规范，确定必需的抽头数量和抽头系数的最简单方法是使用递归FIR滤波器设计程序，例如Remez或者Meteor。

由于按照惯例音频峰值计已经采用了隔直流，在峰值计中去除任何输入直流的影响可能也是一条要求。另一方面，如果为了消除过载而关注真正的峰值信号值，则必须保留并测量直流含量。如果需要，可以在峰值计的输入端引入一个低阶IIR高通滤波器，从而以较低的计算强度去除直流。

为了强调某一部分频带的影响，有时需要在采用了一些类型的加权滤波器以后计量峰值信号幅度，实现方式取决于特定加权滤波器的性质。

附件3

高级音响系统扬声器配置的响度测量算法扩展

1 高级音响系统扬声器配置的扩展

本部分详述了高级音响系统的随意摆放的扬声器配置的客观响度测量算法。

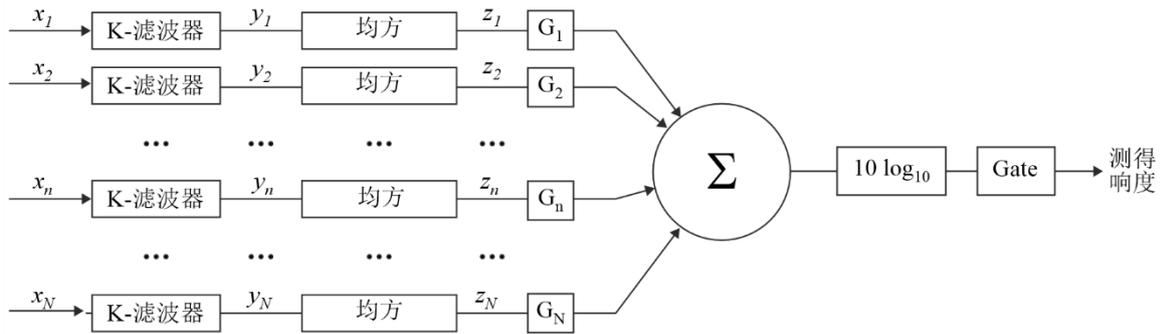
该算法是附件1中规定的3/2多声道音响系统的基本算法的扩展，在该算法中，输入通路的数量增加了，且基本算法的第三个阶段修改如下：

- 信道加权和（除LFE信道以外，每个信道都有一个加权系数 G_i ， G_i 取决于信道位置的方位角和仰角）。

图15呈现了ITU-R BS.2051建议书中详述的高级音响系统扬声器配置客观响度测量算法简略方框图。 N 为输入信道（除LFE信道）的数量。算法的第一阶段、第二阶段和第四阶段（过滤与选通过程）与3/2多声道格式（独立于信道位置）算法中的相同。

图15

高级音响系统扬声器配置客观响度测量算法简略方框图



BS.1770-15

表4给出了信道位置的加权系数 G_i 。 G_i 取决于信道位置的方向，及方位角（ θ ）和仰角（ φ ）。

表4

信道的基于位置的权重

仰角 (φ)	方位角 (θ)		
	$ \theta < 60^\circ$	$60^\circ \leq \theta \leq 120^\circ$	$120^\circ < \theta \leq 180^\circ$
$ \varphi < 30^\circ$	1.00 (± 0 dB)	1.41 (+1.5 dB)	1.00 (± 0 dB)
其他	1.00 (± 0 dB)		

根据表4，在表5中对ITU-R BS.2051建议书中规定的扬声器配置信道的基于位置的权重进行了定义。

表5

ITU-R BS.2051建议书中规定的扬声器配置信道的基于位置的权重

扬声器标签	加权		扬声器配置										
			A	B	C	D	E	F	G	H	I	J	
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0	
M+000	1.00	(±0.0 dB)		X	X	X	X	X	X	X	X	X	X
M+SC	1.00	(±0.0 dB)							X				
M-SC	1.00	(±0.0 dB)							X				
M+030	1.00	(±0.0 dB)	X	X	X	X	X	X	X	X	X	X	X
M-030	1.00	(±0.0 dB)	X	X	X	X	X	X	X	X	X	X	X
M+060	1.41	(+1.5 dB)								X			
M-060	1.41	(+1.5 dB)								X			
M+090	1.41	(+1.5 dB)						X	X	X	X	X	X
M-090	1.41	(+1.5 dB)						X	X	X	X	X	X
M+110	1.41	(+1.5 dB)		X	X	X	X						
M-110	1.41	(+1.5 dB)		X	X	X	X						
M+135	1.00	(±0.0 dB)						X	X	X	X	X	X
M-135	1.00	(±0.0 dB)						X	X	X	X	X	X
M+180	1.00	(±0.0 dB)								X			
U+000	1.00	(±0.0 dB)								X			
U+030	1.00	(±0.0 dB)			X	X	X						
U-030	1.00	(±0.0 dB)			X	X	X						
U+045	1.00	(±0.0 dB)						X	X	X			X
U-045	1.00	(±0.0 dB)						X	X	X			X

表5 (结束)

扬声器标签	加权		扬声器配置										
			A	B	C	D	E	F	G	H	I	J	
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0	
U+090	1.00	(±0.0 dB)									X		
U-090	1.00	(±0.0 dB)									X		
U+110	1.00	(±0.0 dB)				X	X						
U-110	1.00	(±0.0 dB)				X	X						
U+135	1.00	(±0.0 dB)							X	X			X
U-135	1.00	(±0.0 dB)							X	X			X
U+180	1.00	(±0.0 dB)						X		X			
T+000	1.00	(±0.0 dB)								X			
B+000	1.00	(±0.0 dB)					X			X			
B+045	1.00	(±0.0 dB)								X			
B-045	1.00	(±0.0 dB)								X			

附件4

基于对象的音频信号或基于声道和对象的
音频信号组合的响度测量算法

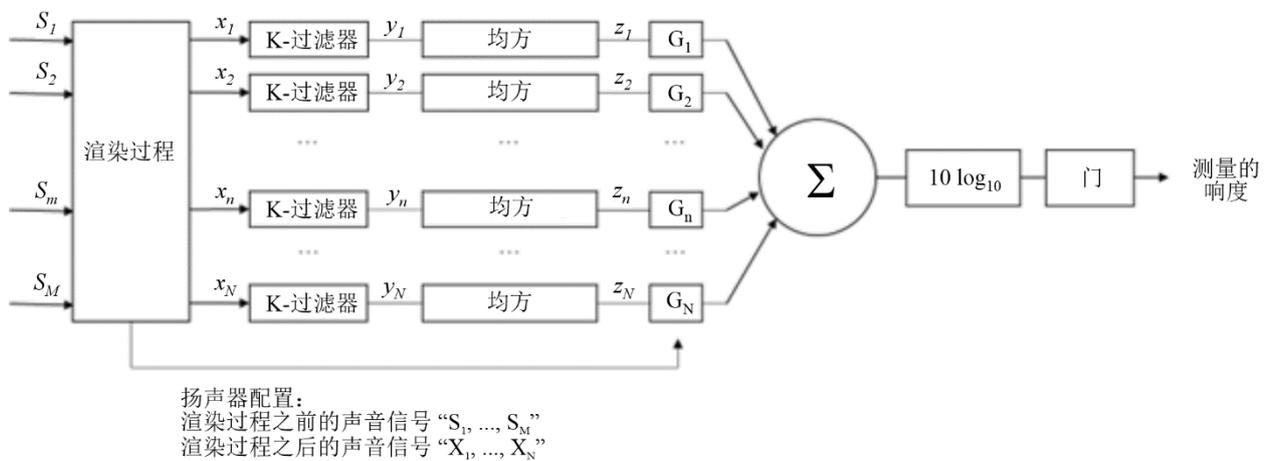
本附件规定了基于对象的音频信号或基于声道和对象的音频信号组合（包括高级音响系统的音频信号）的客观响度测量算法。

基于对象的音频信号或基于声道和对象的音频信号的组合应首先被渲染至特定的扬声器配置。对于高级音响系统，应使用ITU-R BS.2051建议书中规定的扬声器配置。客观响度是使用本建议书附件1和附件3中规定的响度测量算法，从渲染过程输出的音频信号中测量出来的。图16为整个响度测量过程的简化图。

客观响度可能因渲染条件而异。因此，应报告测量中使用的扬声器配置和渲染算法。

图16

基于对象的音频信号或基于声道和对象的音频信号的组合的客观响度测量算法的简化框图



附件4
后附资料1
(资料性)

**基于对象的音频信号或基于声道和对象的音频
信号组合在不同渲染条件下的客观响度**

本附件描述了基于声道和基于对象的音频信号组合在不同渲染条件下的客观响度差异，以及主观测试的结果[3]。

1 不同渲染条件下客观响度的差异

在总共八个节目中，多语种对话、增强对话、多视角对话和背景声音等多个音频对象被制作成完整音频节目的原始声音素材。原始声音素材使用极坐标系统的位置元数据和ITU-R BS.2051建议书中规定的系统H（9+10+3）监测声音系统创建。原始的完整声音素材进行了调整，当使用具有ITU-R BS.2127建议书规定的极坐标的ITU-R ADM渲染器（IAR）渲染到系统H的扬声器配置时，客观响度为-24.0 LKFS。

通过从原始声音素材中选取12-16秒的片段，创建了31种测试素材。剪辑的笛卡尔坐标版本是使用ITU-R BS.2127建议书中规定的转换算法创建的。使用五种渲染算法（极坐标系IAR（IAR-Polar）、具有笛卡尔坐标系的IAR（IAR-Cartesian）、渲染器A和具有默认修剪和无修剪的渲染器B）将一份测试素材渲染到六种扬声器配置（音响系统A（0+2+0）、B（0+5+0）、I（0+7+0）、D（4+5+0）、J（4+7+0）和H（9+10+3））。

注：渲染器B不支持音响系统H（9+10+3）。具有默认修剪和无修剪的渲染器B的行为方式与音响系统J（4+7+0）的行为方式相同。

表6显示了不同渲染条件下客观响度的差异。取决于扬声器配置的客观响度差异，在使用单一渲染算法时可高达2.1至4.1 LU。另一方面，取决于渲染算法的客观响度差异在使用单一扬声器配置时可高达0.3（针对9+10+3）到5.9（针对0+2+0，立体声）LU。当扬声器数量小于音频信号数量时，客观响度会受到渲染算法的影响。因此，应报告测量中使用的渲染条件下的客观响度。

表6

不同渲染条件下客观响度的差异

渲染算法	最大值（绝对值）/平均值（绝对值）	响度配置	最大值（绝对值）/平均值（绝对值），不包括有修剪或无修剪的渲染器B
IAR-Polar	3.1 / 1.1 LU	0+2+0	5.9 / 1.6 LU, 3.6 / 1.9 LU
IAR-Cartesian	3.1 / 1.1 LU	0+5+0	3.9 / 1.0 LU, 2.9 / 1.7 LU
渲染器A	2.1 / 0.6 LU	0+7+0	2.3 / 0.7 LU, 1.7 / 0.9 LU
有修剪的渲染器B	4.1 / 1.9 LU	4+5+0	1.9 / 0.6 LU, 1.7 / 0.9 LU
无修剪的渲染器B	3.8 / 1.0 LU	4+7+0	1.7 / 0.5 LU, 1.7 / 0.5 LU
		9+10+3 ⁽¹⁾	0.3 / 0.0 LU, 0.3 / 0.0 LU

⁽¹⁾ 9+10+3的结果不包括有修剪和无修剪的渲染器B的数据。

2 主观测试

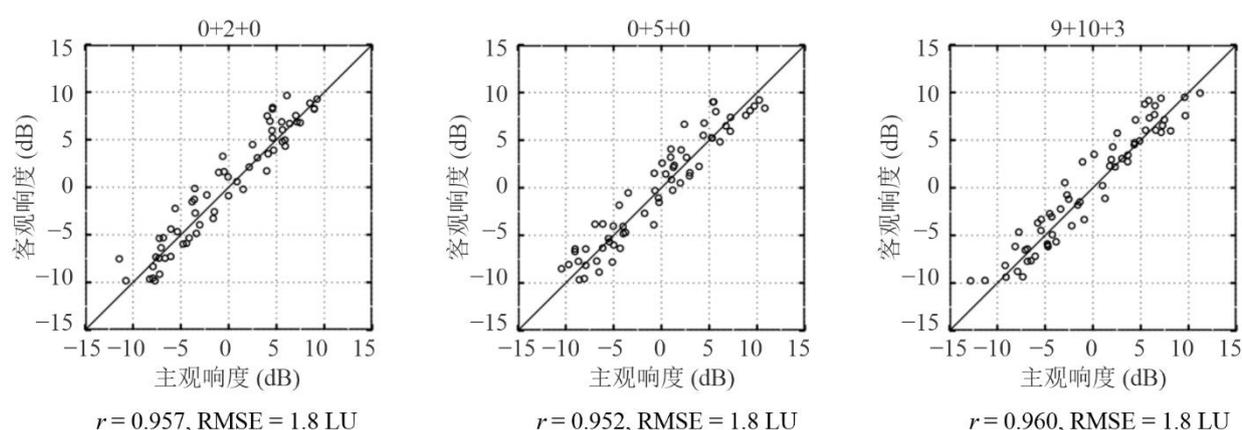
为了评估基于声道和对象的音频信号的响度测量算法，进行了主观测试。主观测试的设备和程序是基于以前的主观测试（见附件1）。

2.1 对测试材料的影响

为了研究对测试材料的影响，测试1中使用了IAR-Polar对三种扬声器配置（0+2+0、0+5+0和9+10+3）渲染的31个测试材料。图17显示了拟议响度算法在两组初始值不同的31个测试材料中的性能。与基于声道的音频信号[1]的结果相比，即使在响度测量中使用渲染处理后的一组音频信号，也不会对结果产生明显影响。

图17

使用三种扬声器配置的31种测试材料的结果

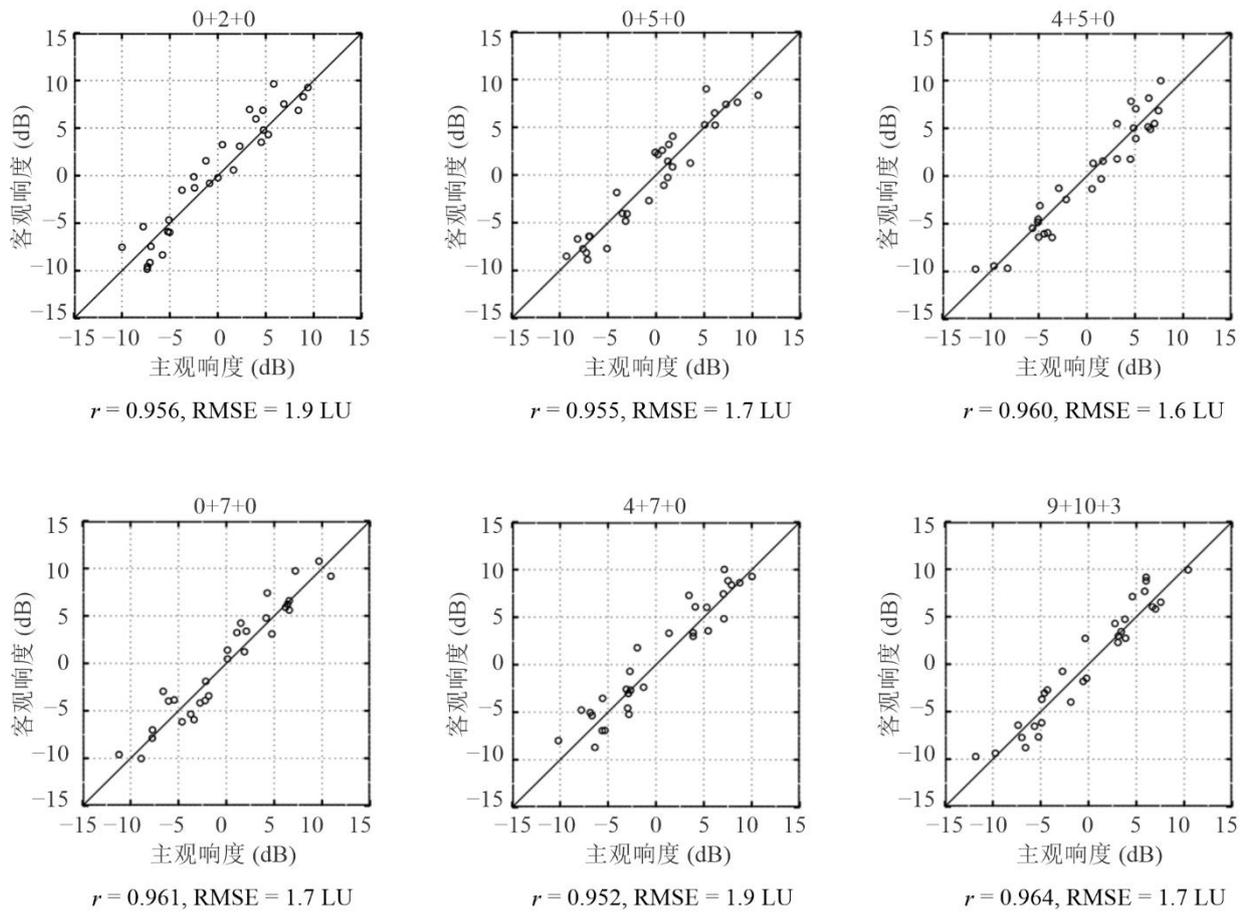


BS.1770-17

2.2 对扬声器配置的影响

为了研究对扬声器配置的影响，测试2中使用了IAR-Polar对六种扬声器配置（0+2+0、0+5+0、4+5+0、0+7+0、4+7+0和9+10+3）渲染的15个测试材料。图18显示了拟议响度算法在六种扬声器配置下的性能。从不同扬声器配置的结果比较来看，对结果没有明显影响。另一项使用四种扬声器配置（0+5+0、4+5+0、0+7+0、4+7+0）的实验也显示了相同的趋势[2]。

图18
六种扬声器配置的结果

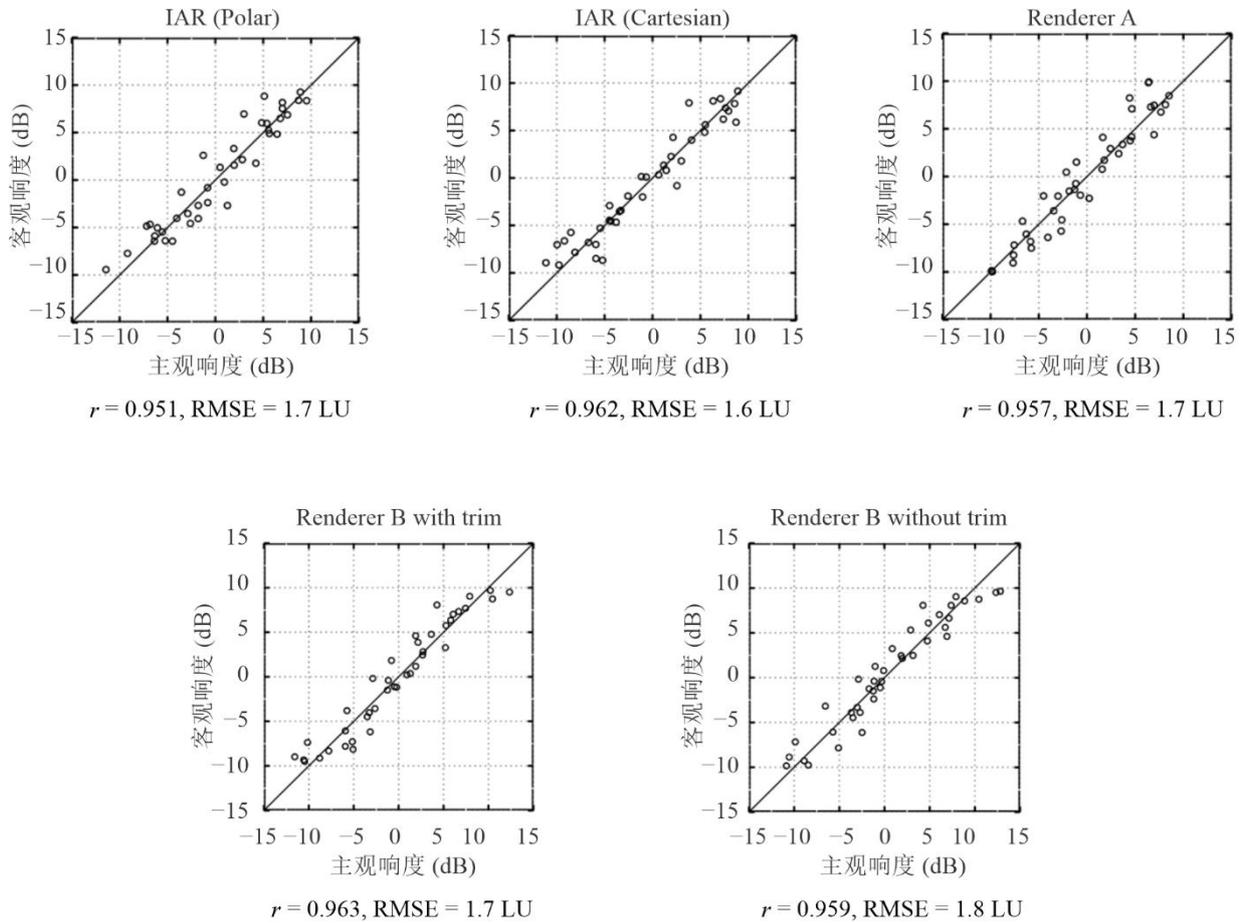


BS.1770-18

2.3 对渲染算法的影响

为了研究对渲染算法的影响，测试3使用了五种渲染算法（IAR-Polar、IAR-Cartesian、渲染器A和有修剪及无修剪的渲染器B）对四种扬声器配置（0+2+0、0+5+0、4+5+0和4+7+0）进行渲染的五个测试材料。图19显示了拟议响度算法在五种渲染算法下的性能。从不同渲染算法的结果比较来看，虽然客观响度因渲染条件不同而不同，但对结果没有显著影响。

图19
五种渲染算法的结果



BS.1770-19

参考文献

- [1] Sizle A.、Sporer T.、Liebetrau J.和Oode S. (2015年), “ITU-R BS.1770建议书中3D音频响度的标准化进展”, 《第三届空间音频国际会议论文集》, 第013号论文。
- [2] Norcross S.、Nanda S.、Cohen Z. (2016年), “基于ITU-R BS.1770的沉浸式音频响度”, 音频工程学会第140届大会, 第9500号大会论文。
- [3] Iwasaki T.、Kubo H.和Oode S. (2022年), “下一代音频内容的响度取决于渲染条件”, 音频工程学会第154届大会, 第77号大会快报。