

ICS 号: xxxx-xxxx
中国标准文献分类号

世界超高清视频产业联盟标准

T/UWA 009.2-4-2026

三维声技术规范 第 2-4 部分：应用指南 响度测量

3D Audio Technology Specification: Part 2-4: Application Guide Loudness

Measurement

（征求意见稿）

xxxx-xx-xx 发布

xxxx-xx-xx 实施

世界超高清视频产业联盟 发布

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 三维声响度测量方法	1
5.1 概述	1
5.2 针对立体声、标准环绕声扬声器配置的节目源的测量方法	1
5.3 针对高阶音响系统扬声器配置的节目源的测量方法	1
5.4 针对包含对象的节目源的测量方法	1
附 录 A （资料性） AudioVivid 响度元数据配置和使用	3
附 录 B （资料性） VBAP 渲染算法说明	4
附 录 C （资料性） VBAP 算法对比测试结果	6
参 考 文 献	9

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》给出的规则起草。

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

本文件由世界超高清视频产业联盟提出并归口。

本文件主要起草单位：华为终端有限公司、中央广播电视总台、广播电视规划院、苏州大学、深圳市腾讯计算机系统有限公司、北京爱奇艺科技有限公司、中国传媒大学、北京数码视讯科技股份有限公司、深圳市奥拓电子股份有限公司、湖南快乐阳光互动娱乐传媒有限公司、中国电子技术标准化研究院。

本文件主要起草人：李正兵、夏丙寅、许凡、陈晨、张建东、汪芮、刘汉源、朱梦尧、李大龙、杜正中、王志航、刘长滔、周骋、梁锋、陈小波、戚成杰、陈仁伟、赵晓莺。

三维声技术规范 第2-4部分：应用指南 响度测量

1 范围

本文件规定了支持菁彩声 Audio Vivid 的三维声节目内容响度测量方法。

本文件适用于符合 Audio Vivid 标准的单声道、立体声、环绕声、包含对象的三维声信号格式的节目收录、制作、交换、传送和检测环节的响度测量，也适用于支持菁彩声 Audio Vivid 节目播放的终端设备（包括但不限于播放器、机顶盒、电视机、投影仪、功放等）的响度控制。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GY/T 262—2012 节目响度和真峰值音频电平测量算法

GB/T 46271—2025 信息技术 三维声技术 编码、分发与呈现

ITU-R BS.1770-5 音频节目响度和真实峰值电平的测量算法 (Algorithms to measure audio programme loudness and true-peak audio level)

3 术语和定义

本文件没有需要界定的术语和定义。

4 缩略语

下列缩略语适用于本文件。

LU: 响度（相对标度）单位（Loudness Unit）

LKFS: K 加权下相对于标称满刻度的响度（Loudness, K-weighted, relative to nominal Full Scale）

RMSE: 均方根误差（Root Mean Square Error）

VBAP: 基于矢量的幅度平移（Vector Based Amplitude Panning）

5 三维声响度测量方法

5.1 概述

本章规定符合菁彩声 Audio Vivid 标准的节目内容的响度测量方法。

5.2 节规定单声道、立体声、标准环绕声扬声器配置的节目源的响度测量方法，5.3 节规定高阶音响系统扬声器配置的节目源的响度测量方法，5.4 节规定包含对象的三维声节目源的响度测量方法。

5.2 单声道、立体声、标准环绕声扬声器配置的节目源的测量方法

应符合 GY/T 262-2012 第3章规定的测量方法。

5.3 高阶音响系统扬声器配置的节目源的测量方法

应符合 ITU-R BS.1770-5 附件3规定的测量方法。

5.4 包含对象的节目源的测量方法

应符合 ITU-R BS.1770-5 附件4规定的测量算法，示意图见图1。

对包含对象信号的三维声（基于对象的音频信号或基于多声道声床和对象组合的音频信号）节目，

按照下列步骤进行测量：

a) 使用对象预渲染算法，将对象预渲染到多声道声床信号上，宜通过 VBAP 算法完成对象预渲染，VBAP 算法定义见附录 B；

b) 按照 5.2 或 5.3 的方法进行响度测量方法。

响度测量结果可能因 VBAP 渲染算法的实现方式不同而有所差异，渲染条件的不同而带来的响度测量差异情况分析见附录 C。

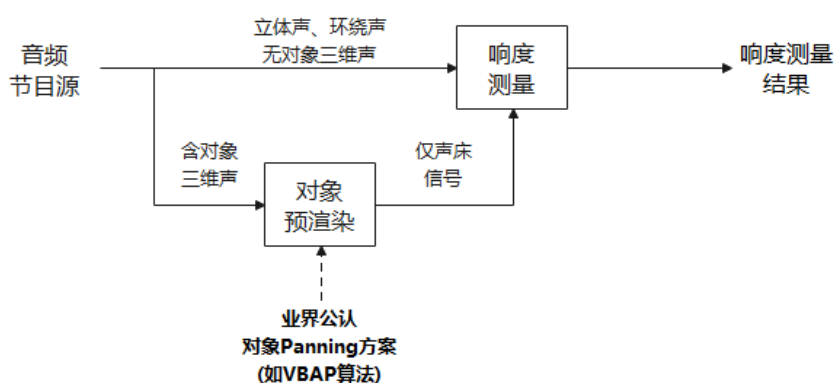


图 1 三维声节目响度测量方案示意图

6 Audio Vivid 响度元数据配置和使用

三维声节目响度元数据可在 Audio Vivid 音频编码过程中输入，并编码到 Audio Vivid 码流中，Audio Vivid 响度元数据定义见参考附录 A。

在三维声节目播放端，可根据从码流中解码得到的响度信息和用户配置的回放音量信息，调节三维声回放的响度水平。

附录 A

(资料性)

Audio Vivid 响度元数据配置和使用

Audio Vivid 响度元数据的属性描述见表 A. 1。

表 A. 1 Audio Vivid 响度元数据参数表

属性名称	含义	取值范围
integratedLoudness	聚合响度值	[-70,0]
loudnessRange	响度范围	[10,70]
maxTruePeak	最大真实峰值	[-70,0]
maxMomentary	最大瞬时响度	[-70,0]
maxShortTerm	最大短期响度	[-70,0]
dialogueLoudness	平均对话响度	[-70,0]

GB/T 46271—2025 9.5 节和 9.6 节规定静态元数据的节目层 (AudioProgramme) 和内容层 (AudioContent) 中宜包含响度元数据。

附录 B

(资料性)

VBAP 渲染算法

B.1 幅度平移的含义

如图B. 1所示，立体声幅度平移（Amplitude panning）可以在人耳形成幻象声源（Phantom），幻象声源的定位（图中虚线）符合下面公式。

$$\sin\theta = \frac{A_L - A_R}{A_L + A_R} \sin\theta_0 \quad \dots\dots\dots \text{B-1}$$

其中， A_L 为左声道幅度， A_R 为右声道幅度。 θ_0 为左右声道与正前方的夹角。所以，幻象声源呈现的位置与左右声道的幅度相关。需要说明的是，上式主要基于双耳ITD（Interaural Time Difference）的推导得到。因此，幻象声源的定位针对700 Hz的低频主要由ITD主导。而随频率升高，双耳定位线索一般认为是ILD（Interaural Level Differences）主导，且随频率升高会有相应变化。

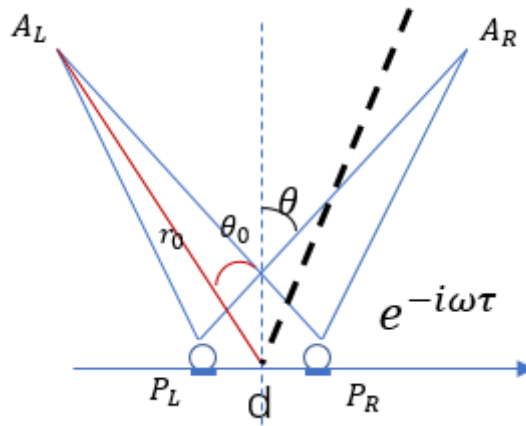


图 B. 1 立体声幅度平移方法

B.2 VBAP的含义

VBAP可以视为立体声幅度平移方法的多通道推广，可由图B. 2所示。其幻象声源的定位公式通常表示为矢量形式，

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2 \quad \dots\dots\dots \text{B-2}$$

其中， $\mathbf{p}$ 是幻象的方位，通常表示为矢量 $[p_x \ p_y]$ 。 $\mathbf{l}_1$ 和 $\mathbf{l}_2$ 是两个声道摆位的向量形式。 g_1 和 g_2 是对应声道的增益系数。显然，针对多声道上式可以进一步表达为矩阵相乘的形式。

$$\mathbf{p} = \mathbf{g} \mathbf{L} \quad \dots\dots\dots \text{B-3}$$

其中， $\mathbf{L}$ 为声道的矩阵形式 $[\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]^T$ 。因此，在二维平面上，VBAP等价于通道1和通道2的立体声幅度平移。在三维空间中，VBAP增加通道3后，幻象声源可以增加高度信息。

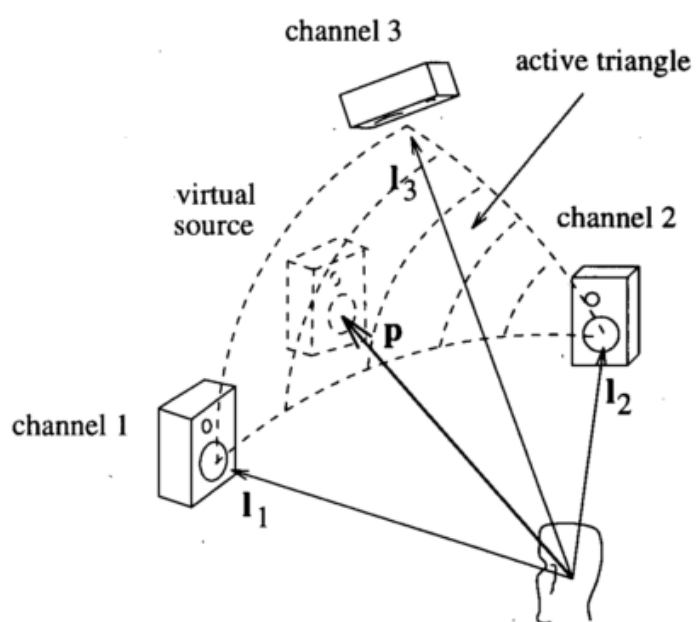


图 B. 2 VBAP 幻象声源定位方法（来源：参考文献 1 Fig. 5.）

B. 3 VBAP渲染算法

通常 VBAP 的渲染过程可以归纳为以下几步：

- 获取扬声器的布局信息，以及待渲染声源（幻象声源）的方位信息；
- 寻找幻象声源周围最接近的扬声器。一般情况下，待渲染对象可以由三只扬声器构成的三角形包围；
- 使用公式B-3得到相应的幅度增益；
- 对于移动声源，声源方位的变化需要重新计算幅度增益；
- 确保幅度增益应用时经过平滑或者相应处理。

VBAP 渲染算法依赖于规则的密集的扬声器布局。同时，VBAP 算法的效果对于环境混响也有依赖。

附录 C

(资料性)

VBAP 算法对比测试结果

C.1 测试设置

测试语料以-24LKFS 为基准（粉噪音源）测试语料库包含 10 首音乐素材。从该语料中顺序提取测试所需的人声音源和音乐音源，人声音源取 10 段，音乐音源取 20 段。

A、B 算法为 VBAP 算法的不同实现。

对于经 A、B 算法处理后的人声音源，每段受试者为 21 人。

10 段人声音源对应的数字域响度设置如下：

- A 算法[-16.5, -17.75, -19, -20.25, -21.5, -24, -26.5, -27.75, -29, -31.5]LKFS
- B 算法[-31.5, -29, -27.75, -26.5, -24, -21.5, -20.25, -19, -17.75, -16.5]LKFS

对于经 A、B 算法处理后的音乐音源，前 10 段受试者为 3 人，后 10 段受试者为 7 人。

前 10 段音乐音源对应的数字域响度设置：

- A 算法[-16.5, -17.75, -19, -20.25, -21.5, -24, -26.5, -27.75, -29, -31.5]LKFS
- B 算法[-31.5, -29, -27.75, -26.5, -24, -21.5, -20.25, -19, -17.75, -16.5]LKFS

后 10 段音乐音源对应的数字域响度设置：

- A 算法[-17, -18.25, -19.5, -20.75, -22, -25, -26, -27.25, -28.5, -30]LKFS
- B 算法[-30, -28.5, -27.25, -26, -25, -22, -20.75, -19.5, -18.25, -17]LKFS

C.2 测试结果

实验过程中，受试者将经 A 算法和 B 算法处理过的同一组语料与粉噪音源响度做对比，并校准其响度使待测语料的响度达到和粉噪音源的听感一致。其中，21 位受试者先对人声音源进行校准，3 位受试者对音乐音源的前 10 段进行了校准，7 位受试者对音乐音源的后 10 段进行校准。

所有音源（包含人声音源和音乐音源）测试结果如图 C.1 所示：

- A 算法下皮尔逊相关系数为 0.97，B 算法下皮尔逊相关系数为 0.98，两者相差 0.01。
- A 算法下 RMSE 值为 2.56LU，B 算法下 RMSE 为 2.36LU，两者相差 0.2LU。

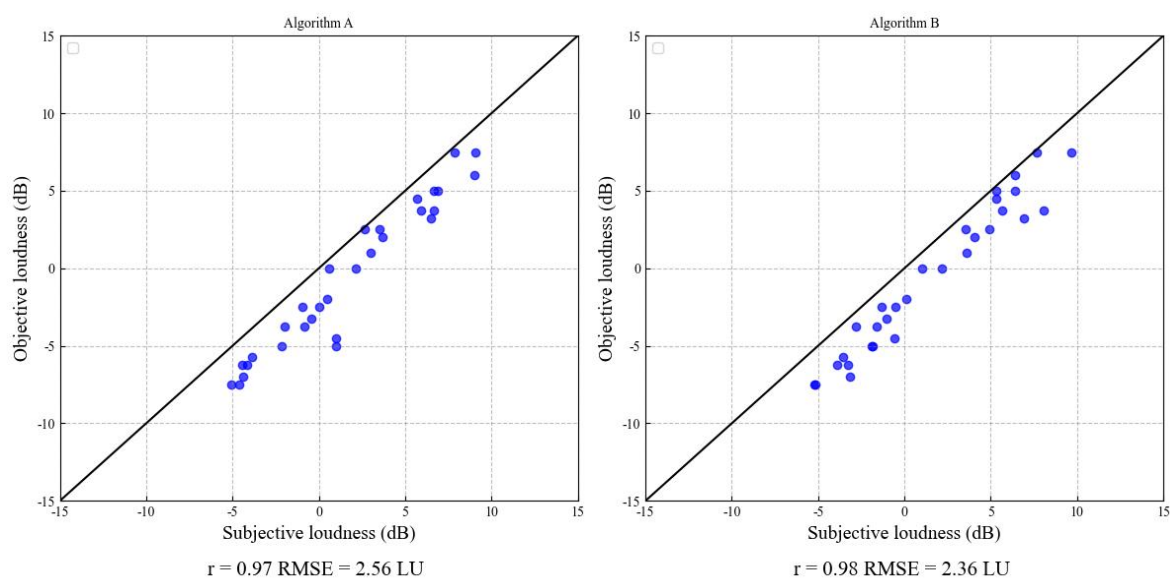


图 C. 1 所有音源测试结果

人声音源测试结果如图 C. 2 所示：

- A 算法与 B 算法下皮尔逊相关系数均为 0.99。
- A 算法下 RMSE 值为 1.85LU，B 算法下 RMSE 值为 1.89LU，两者相差 0.04LU。

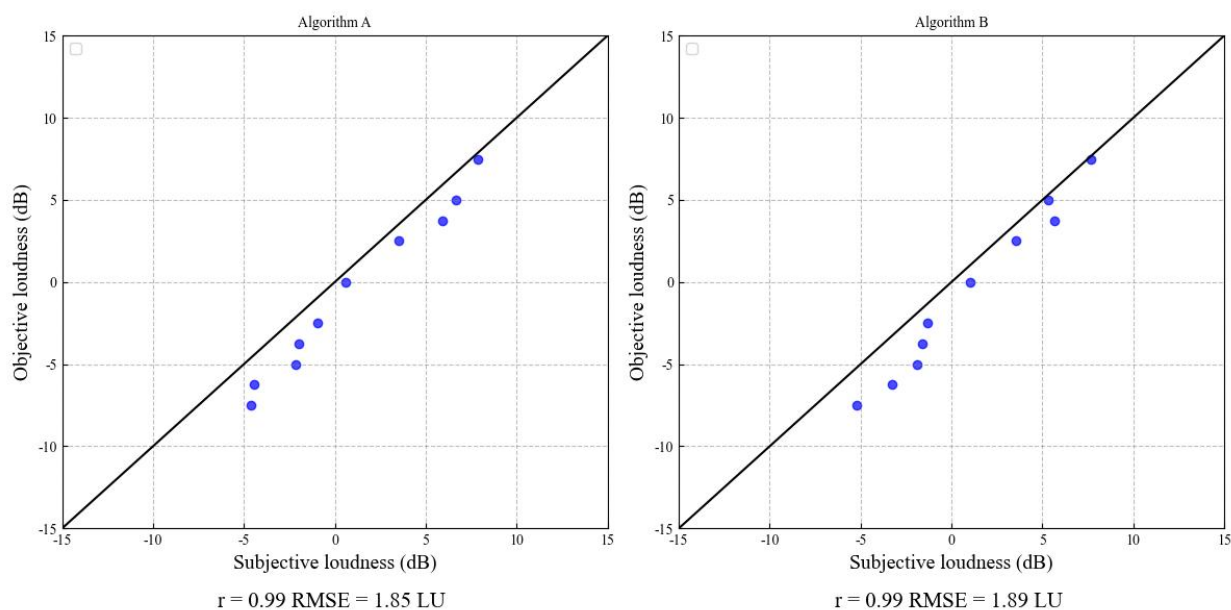


图 C. 2 人声音源测试结果

音乐音源测试结果如图 C. 3 所示：

- A 算法下皮尔逊相关系数值为 0.96，B 算法下皮尔逊相关系数值为 0.98，两者相差 0.02。
- A 算法下 RMSE 值为 2.85LU，B 算法下 RMSE 值为 2.56LU，两者相差 0.27LU。

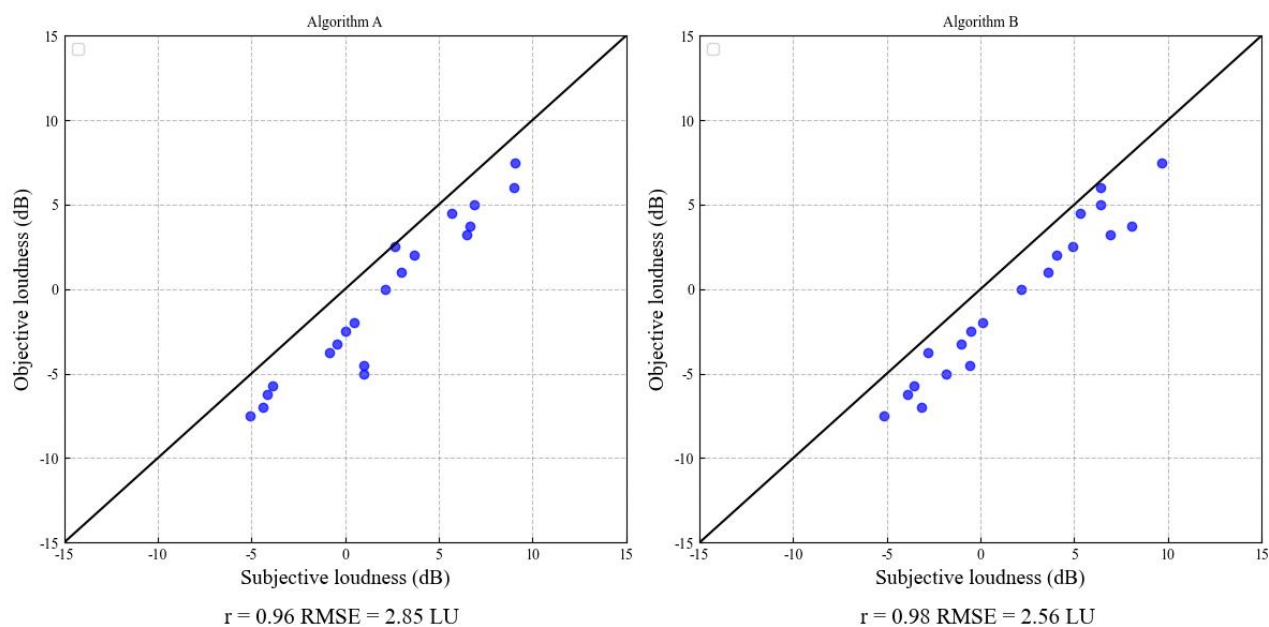


图 C. 3 音乐音源测试结果

参 考 文 献

- [1] VILLE PULKKI, Virtual Sound Source Positioning Using Vector Base Amplitude Panning, 1997.
 - [2] EBU TECH 3341-2023, LOUDNESS METERING: ‘EBU MODE’ METERING TO SUPPLEMENT EBU R 128 LOUDNESS NORMALIZATION.
 - [3] EBU TECH 3342-2023, LOUDNESS RANGE: A MEASURE TO SUPPLEMENT EBU R 128 LOUDNESS NORMALIZATION.
 - [4] EBU R 128 s4-2023, LOUDNESS NORMALISATION OF CINEMATIC CONTENT.
-