

YD

中华人民共和国通信行业标准

YD/T 3245.1—2017

---

远程呈现视频会议系统协议技术要求  
第1部分：媒体参数

Telepresence conference system——Consultative technical requirement  
Part 1: media parameters

2017-04-12 发布

2017-07-01 实施

---

中华人民共和国工业和信息化部 发布

目次

前言.....II

1 范围..... 1

2 规范性引用文件..... 1

3 术语、定义和缩略语..... 1

    3.1 缩略语..... 1

    3.2 术语和定义..... 2

4 范围..... 2

5 远程呈现视频会议端点参数..... 3

    5.1 统一的表示方法..... 3

    5.2 捕获相关参数..... 3

    5.3 环境参数..... 16

6 参数使用方法..... 18

    6.1 捕获区域的使用方法..... 18

    6.2 颜色与光照参数用法..... 21

    6.3 一般参数的使用方法..... 22

    6.4 典型的例子..... 22

## 前 言

YD/T 3245《远程呈现视频会议系统 协议技术要求》分为三个部分：

- 第1部分：媒体参数；
- 第2部分：信令流程；
- 第3部分：媒体传输。

本部分为YD/T 3245的第1部分。

本部分按照GB/T 1.1—2009给出的规则起草。

本部分是“远程呈现视频会议系统”系列标准之一，该系列标准预计的结构及名称如下：

- 《远程呈现视频会议系统 业务需求》；
- 《远程呈现视频会议系统 系统架构》；
- 《远程呈现视频会议系统 协议技术要求》；
- 《远程呈现视频会议系统 服务质量要求》。

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。

本部分由通信标准化协会提出并归口。

本部分起草单位：华为技术有限公司，中国信息通信研究院，中兴通讯股份有限公司

本部分主要起草人：肖 晶、叶小阳、孙明俊、陈 曦。

# 远程呈现视频会议系统协议技术要求

## 第1部分：媒体参数

### 1 范围

本部分确定了远程呈现视频会议系统协议技术要求中的参数部分，主要包括远程呈现视频会议系统的音频、视频和环境参数，并为其提供具体的格式及定义。

本部分适用于远程呈现视频会议系统。

### 2 规范性引用文件

下列文件对于本文件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文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

	远程呈现视频会议系统协议技术要求第2部分：信令流程
IETF CLUE FRAMEWORK	远程呈现系统多码流系统框架
IETF RFC 5646	识别语言标签（Tags for Identifying Languages）
IETF RFC 6351	xCard: vCard XML表示（xCard: vCard XML Representation）
ISO 8995	照明和工作场所（Lighting of work places）
ISO 11664-2	CIE标准光源（CIE standard illuminants）
CIE 15.3:2004	比色法（Colorimetry）
CIE 13.3-1995	光源显色性的测量方法和规定（Method of Measuring and Specifying Colour Rendering Properties of Light Sources）

### 3 术语、定义和缩略语

#### 3.1 缩略语

下列缩略语适用于本文件：

CRI	显色指数	Color Rendition Index
MCU	多点控制单元	Multipoint Control Unit
PC	个人电脑	Personal computer
QoS	业务质量	Quality of Service
SIP	会话初始协议	Session initiation protocol

## 3.2 术语和定义

下列术语和定义适用于本文件：

### 3.2.1

#### **捕获设备 Capture Device**

将音视频输入转换为电信号的设备，通常情况下将馈入媒体编码器。

### 3.2.2

#### **媒体捕获 Media Capture**

媒体的源头，可以来自一个或多个捕获设备，或由其他媒体流构造。

### 3.2.3

#### **远程呈现视频会议系统 Telepresence Conference System**

一系列功能、设备和网络元素的组合，在远程呈现视频会议中，对高质量的交互式音视频信号进行捕获、传输、管理和渲染。通过适当数量的设备（例如摄像头，屏幕，麦克风，扬声器，编解码器，MCU，PC）和特定的环境特征来达到远程呈现的体验。

### 3.2.4

#### **渲染 Render**

从媒体生成表示的过程，例如显示动态视频，或从扬声器发出声音。

## 4 范围

在远程呈现视频会议系统中，需要对多路媒体流进行捕获、编码、传输、解码与呈现。由不同厂商提供的远程呈现视频会议系统，对于类似的问题，可能有不同的解决方案或实现方式。这些解决方案或实现方式，可能使用完全不同的技术，来协商、描述和控制媒体流，这将导致互操作性问题。有些问题可通过翻译或转码来解决，但需要额外的设备，且会引入不必要的延迟并降低性能。

为了防止上述问题，远程呈现会话的能力应可在媒体传输之前就在远程呈现端点之间协商，以商定一致的能力集，若支持相同配置，则采用缺省的能力。

举例而言，接收者可通过能力协商，将其音视频能力参数发送给发送者。发送者将根据接收者的能力进行协调，并将合适的媒体流发送给接收者。

本部分标准将详细描述远程呈现端点的特征及参数，及其如何与远程呈现视频会议系统关联起来。这些参数及其相关联的值可以用来提供高质量的QoS/QoE。

## 5 远程呈现视频会议端点参数

### 5.1 统一的表示方法

本章详细描述远程呈现视频会议系统端点的各种参数。该参数根据如下格式统一定义：

标识	唯一的字符数标识符
描述	该参数的功能描述
格式	参数的类型，如布尔型、字符串、整数、八进制序列、枚举等
可能的取值	该参数可能的取值（范围）
缺省值	可能的缺省值
参考	参考文献
接口	若该参数只适用于部分接口，则需指出
备注	参数定义，或如何使用的补充信息

### 5.2 捕获相关参数

#### 5.2.1 一般参数

##### 5.2.1.1 媒体捕获类型

标识	CaptureMediaType
描述	该参数描述该媒体捕获的类型
格式	枚举
可能的取值	音频，视频，文本...
缺省值	main
参考	IETF CLUEFRAMEWORK中的“CaptureMedia”参数，ITU-T H.TPS-AV中的“Capture MediaType”参数
接口	-
备注	-

##### 5.2.1.2 语言

标识	Language
描述	该参数表示该媒体捕获的内容所使用的语言
格式	枚举
可能的取值	IETF RFC 5646中规定的取值
缺省值	无
参考	IETF CLUEFRAMEWORK中的“Language”参数，ITU-T H.TPS-AV中的“Language”参数
接口	-
备注	-

##### 5.2.1.3 优先级

标识	Priority
描述	该参数表示不同媒体捕获之间的相对优先级

格式	整数
可能的取值	>1
缺省值	无
参考	IETF CLUEFRAMEWORK中的“Priority”参数，ITU-T H.TPS-AV中的“Priority”参数
接口	-
备注	-

5.2.1.4 嵌入文本开关

标识	Embedded text
描述	该参数表示该媒体捕获提供嵌入文本信息
格式	布尔
可能的取值	“是”，该捕获包含嵌入文本； “否”，该捕获不包含嵌入文本
缺省值	否
参考	IETF CLUEFRAMEWORK中的“Embedded text”参数，ITU-T H.TPS-AV中的“Embedded text”参数
接口	-
备注	-

5.2.1.5 捕获相关性

标识	Related to
描述	该参数表示该媒体捕获包含与其它媒体捕获相关的信息
格式	字符串
可能的取值	一个媒体捕获标识
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Related to”参数，ITU-T H.TPS-AV中的“Related to”参数
接口	-
备注	-

5.2.1.6 与会者信息

标识	PersonInfo
描述	该参数允许终端为捕获中的会议与会者提供明确的信息。终端可以自动或手动从不同的源收集信息。使用IETF RFC 6351中的xCard格式来传输信息
格式	字符串
可能的取值	见IETF CLUEFRAMEWORK中的7.1.1.11
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Person Infomation”参数，ITU-T H.TPS-AV中的“Person Infomation”参数。
接口	-
备注	-

## 5.2.1.7 与会者类型

标识	PersonType
描述	该参数表示捕获中的与会者的类型。因一个捕获可以包含多个与会者，故该参数可以包含多个值，但这些值在同一个参数里不会重复
格式	字符串列表
可能的取值	见IETF CLUEFRAMEWORK中的7.1.1.12
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Person Type”参数，ITU-T H.TPS-AV中的“Person Type”参数
接口	-
备注	-

## 5.2.1.8 场景信息

标识	SceneInformation
描述	该参数提供整个捕获场景的信息。终端可以自动或手动从不同的源收集信息。该参数可用来表示的信息包括组织信息或地理信息，以帮助决策和捕获选择。使用IETF RFC 6351中的xCard格式来传输信息
格式	字符串
可能的取值	见IETF CLUEFRAMEWORK中的7.3.1.1
缺省值	-
参考	IETF CLUEFRAMEWORK中的“SceneInformation”参数，ITU-T H.TPS-AV中的“SceneInformation”参数
接口	-
备注	-

## 5.2.1.9 辅流类型

标识	Presentation
描述	该参数表示该媒体捕获源自演示设备。例如，幻灯片、静态图片、数据等等
格式	枚举
可能的取值	Presentation
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Presentation”参数，ITU-T H.TPS-AV中的“Presentation”参数
接口	-
备注	-

## 5.2.1.10 场景描述

标识	Scene Description
描述	该参数使用文字描述整个场景。可以是多种语言的文字
格式	字符串列表
可能的取值	<描述文字>与<语言标签>的列表
缺省值	-
参考	IETF CLUEFRAMEWORK中的“description”参数，ITU-T H.TPS-AV中的“Scene Description”参数

接口	-
备注	使用多种语言描述场景，可帮助用户从多个场景中确定需要的场景

#### 5.2.1.11 场景条目描述

标识	Capture Scene Description
描述	该参数使用文字描述整个场景条目。可以是多种语言的文字
格式	字符串列表
可能的取值	<描述文字>与<语言标签>的列表
缺省值	-
参考	IETF CLUEFRAMEWORK中的“description”参数，ITU-T H.TPS-AV中的“Capture Scene Entry Description”参数
接口	-
备注	使用多种语言描述场景条目，可帮助用户从多个条目中确定需要的场景条目

#### 5.2.1.12 一般空间信息

##### 5.2.1.12.1 捕获描述

标识	Capture Description
描述:	该参数使用文字描述整个场景。可以是多种语言的文字
格式	字符串列表。
可能的取值	<描述文字>与<语言标签>的列表
缺省值	-
参考	IETF CLUEFRAMEWORK中的“description”参数，ITU-T H.TPS-AV中的“Capture Description”参数
接口	-
备注	使用多种语言描述捕获，可帮助用户从多个捕获中确定需要的捕获

##### 5.2.1.12.2 捕获移动性

标识	Mobility of Capture
描述	该参数表示捕获原点、捕获轴点和捕获区域的值是否会改变
格式	枚举。
可能的取值	“静止”，表示捕获原点、捕获轴点和捕获区域的值在会议期间不会改变； “动态”，表示捕获原点、捕获轴点和捕获区域的值在会议期间会改变； “高动态”，表示捕获原点、捕获轴点和捕获区域的值在会议期间会经常改变
缺省值	静止。
参考	IETF CLUEFRAMEWORK中的“Mobility of Capturee”参数，ITU-T H.TPS-AV中的“Mobility of Capture”参数
接口	-
备注	-

## 5.2.1.12.3 视场

标识	View
描述	该参数表示该捕获与哪种类型的视场相关
格式	枚举
可能的取值	“房间”，整个场景； “桌子”，会议桌和与会者； “个人”，一个单独的与会者； “讲台”，教室类型会议中的讲台和演讲者； “听众”，教室类型会议中的听众
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Area Scale”参数，ITU-T H.TPS-AV中的“Area Scale”参数
接口	-
备注	-

## 5.2.1.12.4 场景比例

标识	AreaScale
描述	该参数表示空间信息参数使用的是哪种比例
格式	枚举。
可能的取值	“毫米”，度量系统使用毫米为单位； “未知”，该场景中所有捕获的比例都相同，用于描述相对距离； “无比例”，该场景中每个捕获的比例可能都不同； 其它
缺省值	待决定
参考	IETF CLUEFRAMEWORK中的“Area Scale”参数，ITU-T H.TPS-AV中的“Area Scale”参数
接口	-
备注	-

## 5.2.1.13 一般编码信息

## 5.2.1.13.1 全局最大带宽

标识	maxGroupBandwidth
描述	该参数表示该端点发送的某种特定类型（音频或视频）的所有编码媒体相关的每秒最大比特数
格式	数字
可能的取值	>0
缺省值	无
参考	IETF CLUEFRAMEWORK中的“maxGroupBandwidth”参数，ITU-T H.TPS-AV中的“Overall maximum bandwidth”参数
接口	-
备注	-

5.2.1.13.2 同时联播开关

标识	Simulcast
描述	该参数表示发送方提供同时联播
格式	布尔
可能的取值	“是”，提供同时联播； “否”，不提供同时联播
缺省值	否
参考	ITU-T H.TPS-AV中的“Simulcast”参数
接口	-
备注	-

5.2.2 视频参数

5.2.2.1 视频色域

标识	ColourGamut
描述	该参数表示用于远程呈现视频流中的色域
格式	待决定
可能的取值	BT.709, BT.601, BT.1361
缺省值	BT.709
参考	ITU-T H.TPS-AV中的“Video Colour Gamut”参数
接口	-
备注	若不存在视觉可用性信息（visual usability information, VUI）或其标识为“未指出”时，远程呈现系统需要按照ITU-R BT.709处理所有接收到的视频信号。若传输的视频不符合ITU-R BT.709，发送者应将黑电平、亮度色度范围、传输特性、原色和矩阵系数在视频比特流中传输。接收者可使用这些信息来最优地呈现接收的图像。 具体如何使用颜色与光照参数，请参考第6.2节

5.2.2.2 有效分辨率

标识	EffectiveResolution
描述	该参数表示观看者观察到的呈现后视频流的有效分辨率
格式	浮点数
可能的取值	25~50之间的正数，单位为CPD
缺省值	-
参考	ITU-T H.TPS-AV中的“Effective Resolution”参数
接口	-
备注	-

5.2.2.3 视频捕获数

标识	NumVideoCapture
描述	该参数表示视频捕获的数量
格式	正整数
可能的取值:	>0

缺省值	-
参考	ITU-T H.TPS-AV中的“Number of Video Captures”参数
接口	-
备注	该参数可隐式推导得出。即，所有视频捕获的总数即为该参数的值

5.2.2.4 视频亮度比特深度

标识	LumaBitDepth
描述	该参数表示数字图片中的亮度样本的比特深度
格式	正整数
可能的取值	8~14
缺省值	8
参考	ITU-T H.TPS-AV中的“LumaBitDepth”参数。
接口	-
备注	-

5.2.2.5 视频色度比特深度

标识	ChromaBitDepth
描述	该参数表示数字图片中的色度样本的比特深度
格式	正整数
可能的取值	8~14
缺省值	8
参考	ITU-T H.TPS-AV中的“ChromaBitDepth”参数
接口	-
备注	-

5.2.2.6 视频捕获空间信息

5.2.2.6.1 捕获区域

标识	CapArea
描述	该参数表示该视频捕获在整个捕获场景中的区域
格式	字符串列表。
可能的取值	空间坐标，例如4个（X，Y，Z）的坐标，其中X，Y，Z为整数
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Area of Capture”参数，ITU-T H.TPS-AV中的“Capture Area”参数
接口	-
备注	该参数用于指示视频捕获的区域，亦用于确定视频捕获之间的空间关系。另外，从捕获区域信息还可以得出显示区域之间的空隙宽度（盲区）。在呈现视频流时，呈现者必须考虑该宽度以防止图像畸变

5.2.2.6.2 捕获原点

标识	CapPoint
描述	该参数表示该视频在整个捕获场景中的哪个点捕获
格式	字符串。
可能的取值	空间坐标，例如（X，Y，Z）的坐标，其中X，Y，Z为整数
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Point of Capture”参数，ITU-T H.TPS-AV中的“Capture point”参数
接口	-
备注	-

5.2.2.6.3 捕获轴点

标识	PointonlineofCap
描述	该参数加上捕获原点，共同表示捕获设备的光轴
格式	字符串
可能的取值	空间坐标，例如（X，Y，Z）的坐标，其中X，Y，Z为整数
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Axis of Capture”参数，ITU-T H.TPS-AV中的“Point on line of capture”参数
接口	-
备注	在呈现捕获的时候，该参数可为呈现者提供几何校正信息

5.2.2.7 视频捕获编码信息

5.2.2.7.1 最大视频比特率

标识	maxVideoBitrate
描述	该参数表示单个视频编码的每秒最大比特数
格式	数字
可能的取值	>0
缺省值	无
参考	IETF RFC 6184中的“max-mbps”参数，ITU-T H.TPS-AV中的“Maximum Video Bitrate”参数
接口	-
备注	-

5.2.2.7.2 最大视频分辨率宽度

标识	maxWidth
描述	该参数表示最大视频分辨率宽度，单位为像素
格式	数字
可能的取值	>0
缺省值	无

参考	IETF RFC 6236中的“horizontal image size”参数，ITU-T H.TPS-AV中的“Maximum video resolution width”参数
接口	-
备注	-

#### 5.2.2.7.3 最大视频分辨率高度

标识	maxHeight
描述	该参数表示最大视频分辨率高度，单位为像素
格式	数字
可能的取值	>0
缺省值	无
参考	IETF RFC 6236中的“vertical image size”参数，ITU-T H.TPS-AV中的“Maximum video resolution height”参数
接口	-
备注	-

#### 5.2.2.7.4 最大视频帧率

标识	maxFrameRate
描述	该参数表示最大视频帧率
格式	数字
可能的取值	>0
缺省值	无
参考	IETF RFC 4566中的“framerate”参数，ITU-T H.TPS-AV中的“Maximum video framerate”参数
接口	-
备注	-

### 5.2.3 音频参数

#### 5.2.3.1 音频捕获数

标识	NumAudioCapture
描述	该参数表示音频捕获的数量
格式	正整数
可能的取值	>0
缺省值	-
参考	ITU-T H.TPS-AV中的“Number of Audio Captures”参数
接口	-
备注	该参数可隐式推导得出。即，所有音频捕获的总数即为该参数的值

#### 5.2.3.2 音频捕获空间信息

注：下述参数与视频捕获空间信息中的“捕获区域”和“捕获原点”相同。

5.2.3.2.1 捕获区域

标识	CapArea
描述	该参数表示该音频捕获在整个捕获场景中的区域
格式	字符串列表
可能的取值	空间坐标，例如4个（X，Y，Z）的坐标，其中X，Y，Z为整数
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Area of Capture”参数，ITU-T H.TPS-AV中的“Capture Area”参数
接口	-
备注	-

5.2.3.2.2 捕获原点

标识	CapPoint
描述	该参数表示该音频在整个捕获场景中的哪个点捕获
格式	字符串
可能的取值	空间坐标，例如（X，Y，Z）的坐标，其中X，Y，Z为整数
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Point of Capture”参数，ITU-T H.TPS-AV中的“Capture point”参数
接口	-
备注	-

5.2.3.3 音频捕获编码信息

5.2.3.3.1 音频声道格式

标识	AudioChannelFormat
描述	该参数表示该音频声道的格式
格式	枚举
可能的取值	“单声道”，“立体声”
缺省值	-
参考	IETF CLUE 架构文稿中的“AudioChannelFormat”参数，ITU-T H.TPS-AV中的“AudioChannelFormat”参数
接口	-
备注	-

5.2.3.3.2 最大音频比特率

标识	maxAudioBitrate
描述	该参数表示单个音频编码的每秒最大比特数
格式	数字
可能的取值	>0
缺省值	-

参考	IETF RFC 4566中的“maxBitRate”参数，ITU-T H.TPS-AV中的“Maximum Audio Bitrate”参数
接口	-
备注	-

#### 5.2.3.4 名义音量

标识	nominalAudioLevel
描述	该参数表示远程呈现音频流中的名义音量
格式	数字，单位为dBov
可能的取值	<0
缺省值	-26
参考	ITU-T H.TPS-AV中的“NominalAudio Level”参数
接口	-
备注	-

#### 5.2.3.5 动态音量

标识	dynamicAudioLevel
描述	该参数表示远程呈现音频流中的实际音量
格式	数字，单位为dBov
可能的取值	-127~0
缺省值	-26
参考	ITU-T H.TPS-AV中的“Dynamic Audio Level”参数
接口	-
备注	-

#### 5.2.3.6 发送响度评定

标识	SendLoudnessRating
描述	该参数表示远程呈现音频流中的声到电传输函数
格式	数字，单位为dB
可能的取值	+4~+16
缺省值	(+13-Fs)
参考	ITU-T H.TPS-AV中的“Sending LoudnessRating”参数
接口	-
备注	在声音捕获阶段进行音频对准的目的是： 设置麦克风前置放大器增益值，以在非饱和的情况下使用全动态模数转换器。 为运行中的所有房间的所有麦克风设置前置放大器增益值，使得在模数转换前，在所有与会者的位置，相同的声压产生相同的电信号值

#### 5.2.3.7 接收响度评定

标识	ReceiveLoudnessRating
描述	该参数表示远程呈现音频流中的电到声传输函数

格式	数字，单位为dB
可能的取值	-26~+18
缺省值	(+5-Fr)
参考	ITU-T H.TPS-AV中的“Receiving LoudnessRating”参数
接口	-
备注	在声音呈现阶段进行音频对准的目的是： 设置扬声器增益值，以在与会者的位置获得舒适的音量（既不太高又不太低）。 为麦克风设置扬声器增益值，使得在数模转换后，在所有与会者的位置，来自远端房间的相同的电信号值产生相同的声压

5.2.3.8 加权麦克风扬声器耦合损耗

标识	weightedMicrophoneLoudspeakerCouplingLoss
描述	该参数表示麦克风扬声器对之间的耦合损耗。实际上，最恶劣的情况（即耦合损耗为最小）是麦克风和扬声器摆得很靠近
格式	数字，单位为dB
可能的取值	标称音量≥ 6dB，最大音量≥40dB
缺省值	-
参考	ITU-T H.TPS-AV中的“weightedMicrophone-LoudspeakerCouplingLoss”参数
接口	-
备注	最大化该参数可降低回声抵消器的操作难度，改善系统的全双工性能

5.2.4 延迟参数

5.2.4.1 端到端视频延迟

标识	EndtoEndVideoDelay
描述	该参数表示两个远程呈现视频会议端点之间的视频媒体的单向端到端延迟
格式	整数，单位为ms
可能的取值	-
缺省值	-
参考	ITU-T H.TPS-AV中的“EndtoEndVideoDelay”参数
接口	-
备注	为了给终端用户提供优秀的QoE体验，该参数应小于325ms

5.2.4.2 端到端音频延迟

标识	EndtoEndAudioDelay
描述	该参数表示两个远程呈现视频会议端点之间的音频媒体的单向端到端延迟
格式	整数，单位为ms
可能的取值	-
缺省值	-
参考	ITU-T H.TPS-AV中的“EndtoEndAudioDelay”参数
接口	-
备注	为了给终端用户提供优秀的QoE体验，该参数应小于325ms

### 5.2.4.3 音视频同步

标识	AudioVideoSynchronization
描述	该参数表示视频媒体与对应音频媒体之间的同步差异（端到端视频延迟-端到端音频延迟）
格式	整数，单位为ms
可能的取值	-
缺省值	-
参考	ITU-T H.TPS-AV中的“AudioVideoSynchronization”参数
接口	-
备注	为了给终端用户提供优秀的QoE体验，该参数应处于40ms~60ms之间

## 5.2.5 多内容捕获参数

### 5.2.5.1 多内容捕获开关

标识	multiContentCapture
描述	该参数表示该捕获是否包含多个捕获源
格式	布尔
可能的取值	“是”，表示该捕获包含多个捕获源； “否”，表示该捕获包含单个捕获源
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Multiple content capture”参数，ITU-T H.TPS-AV中的“Multiple Content Capture”参数
接口	-
备注	-

### 5.2.5.2 多内容捕获源

标识	MCC Source
描述	该参数表示该媒体流中包含的多个捕获源的列表。
格式	列表
可能的取值	媒体捕获源的ID列表
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Multiple content capture”参数，ITU-T H.TPS-AV中的“Multiple Content Capture Source”参数
接口	-
备注	-

### 5.2.5.3 最大多内容捕获源

标识	MCC maxCapture
描述	该参数表示任意时刻在一个媒体流中捕获源的最大个数
格式	整数
可能的取值	>1
缺省值	-

参考	IETF CLUEFRAMEWORK中的“Maximum Number of Captures within a MCC”参数，ITU-T H.TPS-AV中的“Maximum Multiple Content Capture Sources”参数
接口	-
备注	-

5.2.5.4 多内容捕获策略

标识	MCC policy
描述	该参数表示媒体流中的捕获源是如何选择的
格式	枚举
可能的取值	见IETF CLUEFRAMEWORK中的7.2.1.2
缺省值	-
参考	IETF CLUEFRAMEWORK中的“Policy”参数，ITU-T H.TPS-AV中的“Multiple Content Capture Policy”参数
接口	-
备注	-

5.2.5.5 多内容捕获同步

标识	MCC synchronizationID
描述	该参数为多个MCC捕获之间提供连接，用以在特定时刻同步出现在媒体流中的捕获
格式	字符串
可能的取值	见IETF CLUEFRAMEWORK中的7.2.1.3
缺省值:	-
参考:	IETF CLUEFRAMEWORK中的“Synchronisation Identity”参数，ITU-T H.TPS-AV中的“Multiple Content Capture Synchronisation”参数
接口:	-
备注:	-

5.3 环境参数

5.3.1 颜色与光照参数

5.3.1.1 概述

要获得令人满意的远程呈现体验，远程呈现端点需配备合适数量和质量的光照。ISO 8995为室内环境的光照设计定义了一系列评判标准。它强调下列参数对发光环境有重要作用：

- 亮度分布；
- 照明；
- 强光；
- 光的方向性；
- 光线与发光面的颜色方面；
- 闪烁；
- 日光；

——保养。

当设计和试车远程呈现端点时，需要考虑上述方面。

### 5.3.1.2 光源

标识	Type
描述	该参数表示麦克风扬声器对之间的耦合损耗。实际上，最恶劣的情况（即耦合损耗为最小）是麦克风和扬声器摆得很靠近
格式	枚举
可能的取值	“Dx” 代表日光（x为整数，D65为默认，其他包括D50,55,75）； “E” 代表等能量散热器； “FLx” 代表日光灯（x为整数，例如FL1，FL2，FL12等）； “FL3.x” 代表三带日光灯（x为整数，例如FL3.1，FL3.14等）； “HPx” 代表高压灯族（x为整数， HP1代表标准高压灯； HP2代表增色高压钠灯； HP3到HP5代表高压金属卤化物灯）
缺省值	-
参考	ITU-T H.TPS-AV中的“Illuminant”参数，b-CIE 15.3:2004以及ISO 11664-2
接口	-
备注	若CRI值和色温值省略，则他们可以从光源类型推导出来

### 5.3.1.3 显色指数

标识	CRI
描述	该参数表示在远程呈现端点的可见（环境）光的显色指数。其计算方法可见b-CIE 13.3-1995
格式	数字
可能的取值	>0
缺省值	-
参考	ITU-T H.TPS-AV中的“Colour Rendering Index”参数
接口	-
备注	若该光源是标准CIE光源，则显色指数可省略

### 5.3.1.4 色温

标识	ColTemp
描述	该参数表示在远程呈现端点的可见（环境）光的相关色温
格式	数字，单位为开尔文
可能的取值	>0
缺省值	-
参考	ITU-T H.TPS-AV中的“Colour temperature”参数
接口	-
备注	若该光源是标准CIE光源，则色温值可省略

5.3.2 混响时间

标识	ReverberationTime
描述	该参数表示在一个远程呈现会议室中的混响时间
格式	数字，单位为s
可能的取值	0.3~0.4
缺省值	-
参考	ITU-T H.TPS-AV中的“ReverberationTime”参数
接口	-
备注	-

5.3.3 环境噪声

标识	AmbientNoise
描述	该参数表示在远程呈现会议室中，在所有相应坐席位置测量的等A加权声压值
格式	数字，单位为dBA
可能的取值	≤30
缺省值	-
参考	ITU-T H.TPS-AV中的“AmbientNoise”参数
接口	-
备注	-

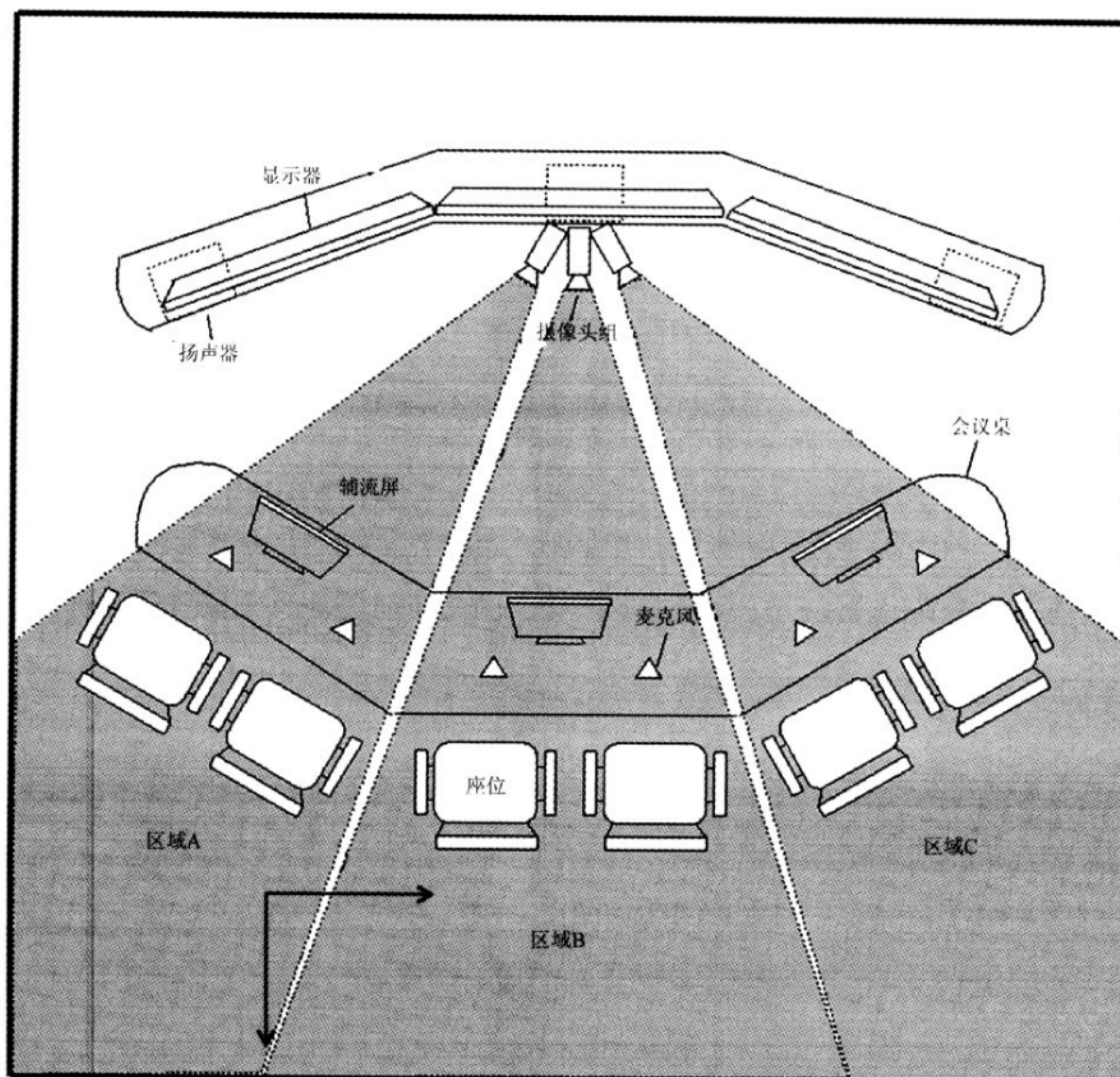
5.3.4 隔音

标识	SoundInsulation
描述	该参数表示加权后的远程呈现会议室与相邻房间之间的声音衰减指数（ $R'_w$ ）
格式	数字，单位为dB
可能的取值	≥48
缺省值	-
参考	ITU-T H.TPS-AV中的“SoundInsulation”参数
接口	-
备注	-

6 参数使用方法

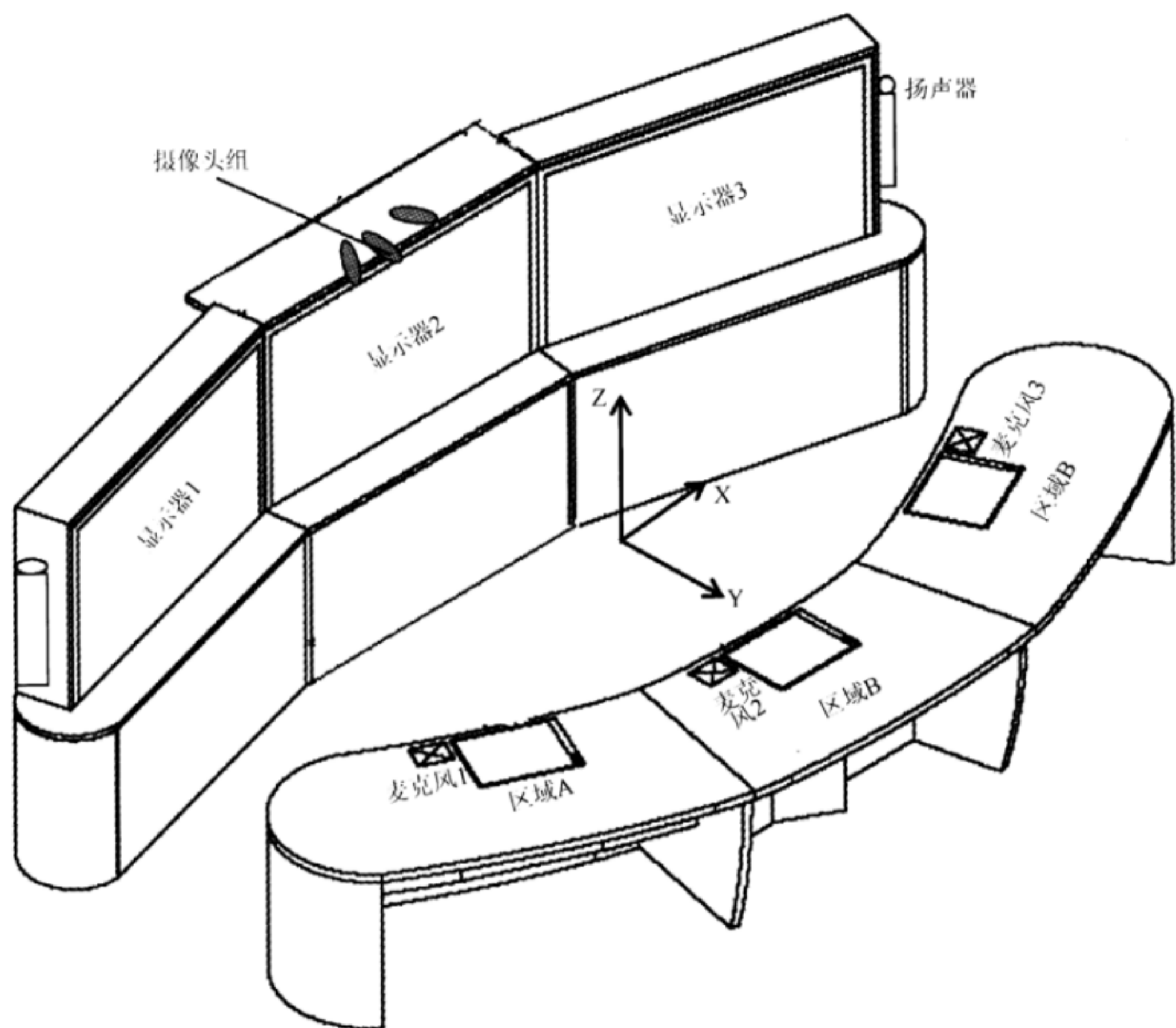
6.1 捕获区域的使用方法

捕获者发送捕获区域和场景区域等参数给呈现者，这些参数通常以空间信息的形式表示，例如四个  $X$ ， $Y$ ， $Z$ 组成的坐标组。举例而言，捕获区域可以由共面的四个点组成，分别为左下（ $X1$ ， $Y1$ ， $Z1$ ）、右下（ $X2$ ， $Y2$ ， $Z2$ ）、左上（ $X3$ ， $Y3$ ， $Z3$ ）、右上（ $X4$ ， $Y4$ ， $Z4$ ）。按照惯例，随着数字从小到大，对应的摄像头拍摄位置从左到右（在 $X$ 方向），观众从前到后（在 $Y$ 方向）或从下到上（在 $Z$ 方向）。俯视图如图1所示，三维图如图2所示。



注：“捕获原点”“捕获轴点”等参数未在图示中呈现。

图1 捕获区域坐标轴（俯视图）



注：“捕获原点”“捕获轴点”等参数未在图示中呈现。

图2 捕获区域坐标轴（三维图）

捕获区域参数可以为视频流提供多方面的信息。为简便起见，下述例子仅以X轴为对象展开讨论，而对Y轴和Z轴同样适用。假定使用mm作为比例单位。

首先，捕获区域可以用来指示视频捕获的空间信息。例如，区域A的捕获区域见表1。

表1 区域A的捕获区域

	左下	右下	左上	右上
区域 A	(0,2000,0)	(980,2200,0)	(0,2000,600)	(980,2200,600)

这样，可以得出视频流的（拍摄）宽度信息，即水平方向（沿着X轴）的980毫米，从而获得调整参数（拍摄宽度信息或拍摄区域之间的盲区）。呈现者可使用调整参数信息，将图像呈现到正确的空间位置上；也可以使用这个信息，来调整图像的尺寸，以符合显示器的最佳显示尺寸；还可以使用这个信息，来调整呈现者要发送的媒体流，使其与区域A的宽度一致。

其次，捕获区域可以用来描述多个视频捕获之间的空间关系。再考虑区域B见表2所示。

表2 区域B的捕获区域

	左下	右下	左上	右上
区域B	(990,2200,0)	(1990,2200,0)	(990,2200,600)	(1990,2200,600)

通过分别分析区域A和区域B的捕获区域，可以得到这两个视频捕获之间的空间关系，即区域A在区域B的左边、区域B在区域A的右边。因此，呈现者可使用这些信息，将图像呈现在正确的空间位置。

再次，捕获区域可以用来确定多个视频捕获之间的盲区。举例而言，通过分析区域A和区域B各自的X坐标（即980与990），可以得出它们之间的盲区为10mm。这样，用空间信息的形式描述盲区，可以明确地指示其相对视频图像的位置关系，呈现者可以根据该信息，准确地对待呈现的视频图像进行调整。同样地，呈现者可使用这个信息，对图像进行调整，使相邻图像之间拼接无错位。即，盲区的宽度为相邻图像之间的水平坐标差。当盲区宽度与呈现者的显示屏之间的缝宽不同时，需要对图像进行调整，以避免相邻图像之间出现错位、扭曲等现象。通常，可以对待显示的图像进行裁剪然后填充、或缩小然后填充、或放大然后裁剪等方法，具体采用哪种方式可以通过计算盲区宽度与缝宽的比例来确定。填充的内容可以为纯色、预设的颜色或者相邻图像的内容。

## 6.2 颜色与光照参数用法

远程呈现的一个目的是提供“身临其境”的体验。人体视觉系统有自适应调节能力，在同一个场景下，即使处于不同的光照环境下，物体仍会呈现相同的颜色。所以，处于特定场景里的人，无论环境光如何改变，看到物体的颜色不会改变。然而，在远程呈现系统中，与会者通常处于不同的场景，光照环境也并非完全一样。因此，远端与会者看到的本地内容将和本地与会者看到的不同。

摄像头等捕获设备可以对捕获的图像执行诸如“白平衡”等图像转换方法，以让其颜色与人眼看到的更接近。要改变图像的颜色，捕获设备应知道环境光（光源）的特征（如白平衡点），以正确地调整图像的颜色。5.3.1中的“光源”“显色指数”和“色温”都是可用于图像转换的本地环境参数。

当图像传到远端并显示时，在多数情况下，它都会处于不同的环境光中。如果远端与会者看到的图像与本地不同，则会缺乏“身临其境”的体验。为了最大化“身临其境”的体验，显示设备需综合考虑本地会场的环境光和远端会场的环境光，以保证所有图像的显示一致。本地会场的环境参数可以基于事先统一的设置，而远端会场的环境参数可以通过同代信号传输。5.2.2.1中的“视频色域”参数是传输这个信息的一个途径，另外也可以通过信令协议携带这些参数。通常，使用反映不同光照参数之间关系的图像转换系数（例如颜色变换矩阵、3D色彩空间插值表），来转换图像，使其为与目标环境一致。

具体来说，远端会场将其图像和光照参数发送给本地会场，本地会场根据远端会场的光照参数，结合本地的光照参数，获得图像转换系数信息；再根据图像转换系数，对待显示的图像进行图像转换，因为转换后图像的光照参数与本地会场的光照参数一致，所以在本地会场显示的时候，转换后的图像可以在本地获得“身临其境”的体验。或者，本地会场根据本地的光照参数，结合远端会场的光照参数，获得图像转换系数信息；再根据图像转换系数，对待发送的图像进行图像转换，转换后的图像将与远端光照环境一致，即在远端获得“身临其境”的体验。通常，图像转换系数通过查找本地和远端的光照参数获得，由颜色变换矩阵或3D颜色空间查找表表示，原图像乘以颜色变换矩阵即可获得目标图像。

6.3 一般参数的使用方法

5.2.1中规定的一般参数，如“优先级”“合成”“自动切换”“场景描述”“场景切换策略”“语言”“演示”“角色”等，从不同角度，在媒体流的内容方面对媒体流进行描述。接收端可以根据这些参数，对媒体流进行区分，从而可以选取自己感兴趣的媒体流。

例如，如果接收端想观看发言者的视频，则其可以根据参数“自动切换”（或“自动切换”与“场景切换策略”），选择对应的视频流。

又例如，当发送端为三屏系统，而接收端为单屏系统时，如果接收端想观看发送端所有的图像，则其可以结合参数“媒体捕获内容”“合成”“场景描述”等参数，选择对应的视频流。

又例如，接收端根据这些参数（或其中的一部分），为每个媒体流分配相应的重要性（优先级）等级（多个则为其加权和），当拥塞发生或拥塞解除时，发送端根据接收端分配的重要性对不同媒体流进行不同的带宽调整。这种方法特别适用于多媒体流复用于同一端口的情况。举例说明，通常发言者的画面是更受关注的，所以可以为“自动切换”（或“自动切换”与“场景切换策略”）分配较高的权值，在拥塞发生时，优先保证其质量不受影响（例如暂停传输重要性等级低于阈值的媒体流，或按照重要性等级等比例降低带宽）。

又例如，对于发送端，它将一个或多个“多捕获内容”和媒体捕获相关的“媒体捕获标识”及其他参数（例如场景描述、场景条目描述，每个场景条目都包括从同一场景采集的媒体捕获或多捕获内容）包含在告知消息中，并发送给接收端；它接收包含接收端根据告知消息而选中的“多捕获内容”或媒体捕获的配置消息；它根据选择结果，向接收端发送对应的媒体数据流。

因此，通过使用这些参数，媒体接收端可以更为了解媒体流的内容，从而辅助发送端进行更符合接收端观看者体验的决策。

6.4 典型的例子

在呼叫建立之后、媒体传输之前，远程呈现视频会议端点与MCU或远程呈现视频会议端点之间，可以通过远程呈现视频会议信令，协商本标准中规定的参数。

详细的协商流程见YD/T 3245.2—2017，这里仅以一个简单的例子介绍如何协商、使用这些参数。

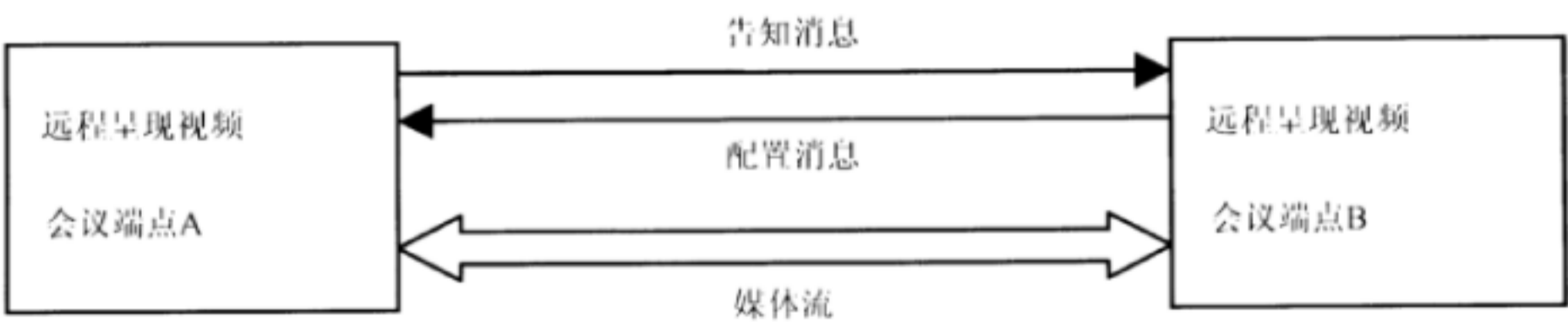


图3 媒体参数协商模型

如图3所示，远程呈现视频会议端点A（以下简称端点A）与远程呈现视频会议端点B（以下简称端点B）通过告知消息和配置消息进行参数协商。具体来说，首先，端点A向端点B发送告知消息。这个告知消息中包含两个或多个端点A能够为端点B提供的媒体捕获，以及相关的能力信息，每条能力信息都包含“媒体捕获标识”和“捕获区域”参数，可以包含“一般参数”、其它“音视频参数”“延迟参数”以及各种“环境参数”。

然后，端点B根据这些能力信息，选择其想接收的媒体捕获，并将选择结果和相应的能力信息，通过配置消息反馈给端点A。

端点A根据配置消息，确定要发送给端点B的媒体捕获，接下来即可建立媒体传输通道，传输相对应的媒体流。

---