



基于客观指标的音视频质量测试的自动化实践

X-Live RTC·进阶实战高手课

目录

- 1 为什么需要质量评估
- 2 音视频质量评估
- 3 音视频客观评价标准
- 4 基于客观指标的自动化测试
- 5 音视频质量评价的挑战与未来

为什么需要质量评估

音视频通话的基本流程



为什么需要质量评估

- 影响质量的环节多，场景复杂，问题定位困难。
- 缺少评价手段和统一标准，没法衡量效果。
- 开发人员和测试人员对质量的理解差异，导致沟通成本高。
- 传统测试，测试流程长，效率低。

目录

- 1 为什么需要质量评估
- 2 **音视频质量评估**
- 3 音视频客观评价标准
- 4 基于客观指标的自动化测试
- 5 音视频质量评价的挑战与未来

音视频质量评估

评估维度：

听感	可懂度、卡顿、回声、噪音
观感	清晰度、流畅度
综合体验	端到端延时、音画同步等
QoS	码率、视频分辨率、音频采样率、CPU、内存、发热耗电

音视频质量评估

评估方法:

1. 主观质量评估

在国际标准中，统一使用 MOS 值来评价音视频的主观体验。MOS 值按主观感受从1-5打分。

MOS分	主观体验
5分	非常好
4分	好
3分	一般
2分	差
1分	不可用

音视频质量评估

主观质量评估一般流程：

- 选取测试音视频
- 衡量音视频质量的方案
- 设计主观评估实验
- 看视频打分（MOS 分）

主观质量评估标准：

- P.800 《传输质量的主观评定办法》
- P.830 《话带和宽带数字语音编码器的主观评价方法》
- BT.500 《电视图像质量的主观评估方法》
- P.910 《多媒体应用的主观性视频质量评价方法》

音视频质量评估

2. 客观质量评估

利用数学模型模拟主观评估结果，输入参数不变，评估出的结果一样。

客观质量评估分类：

- 全参考评估：原始图像/音频和有损图像/音频各个层面比较，输出结果。
- 部分参考评估：原始图像/音频和有损图像/音频中提取一些特征值作比较，输出结果。
- 无参考评估：不需要原始图像/音频，直接使用有损图像/音频的一些特征值，输出结果。

音视频质量评估

主客观质量评估比较：

方法种类	方法说明	方法特点
主观评估	利用人的感官对音视频进行质量评判并得到 MOS 分值。	准确性高 实施成本高 可重复性差 无法大批量评估
客观评估	利用数学模型，通过一些评估标准来量化音视频质量。	准确性依赖于评估的数学模型 可重复性高 可大批量评估

目录

- 1 为什么需要质量评估
- 2 音视频质量评估
- 3 **音视频客观评价标准**
- 4 基于客观指标的自动化测试
- 5 音视频质量评价的挑战与未来

音视频客观评价标准

1. 音频客观评价标准

有参考：

P.862 PESQ、P.863 POLQA 和 VISQOL。

无参考：

P.563 和 G.107 E-Model

2. 视频客观评价指标评价标准

有参考：

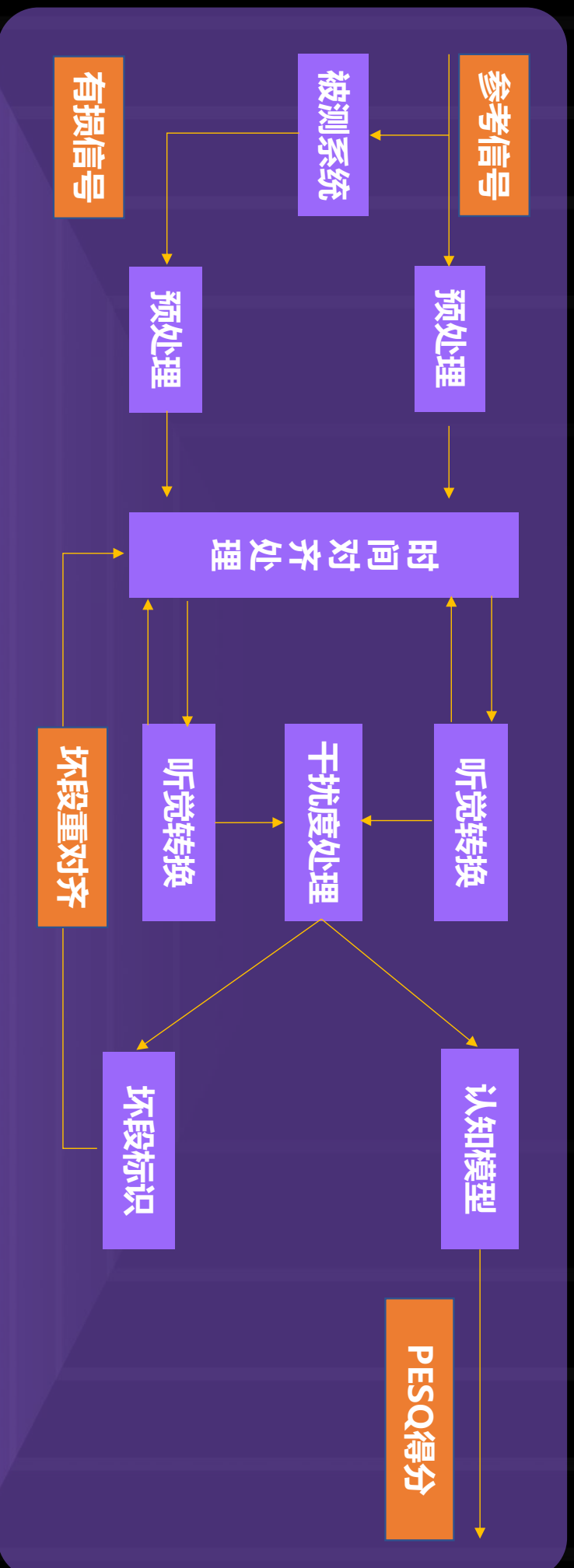
PSNR、SSIM、VMAF

无参考：

BRISQUE、RankIQ、DIOA

音视频客观评价标准

PESQ 算法流程



音视频客观评价标准



音视频客观评价标准

POLQA

	PESQ		POLQA
Codecs	AMR EFR	AMR AMR-WB EFR EVRC EVRC-B EVRC-WB	iLBC AMB+ AAC Skype / SILK G.711 G.729
Reference Speech (sampling frequency)	8 kHz	8 kHz 48 kHz	
Applications	POTS VoIP 3G	HD Voice Voice Enhancement Devices	Skype Calls Benchmarking CDMA and GSM

音视频客观评价标准

PSNR

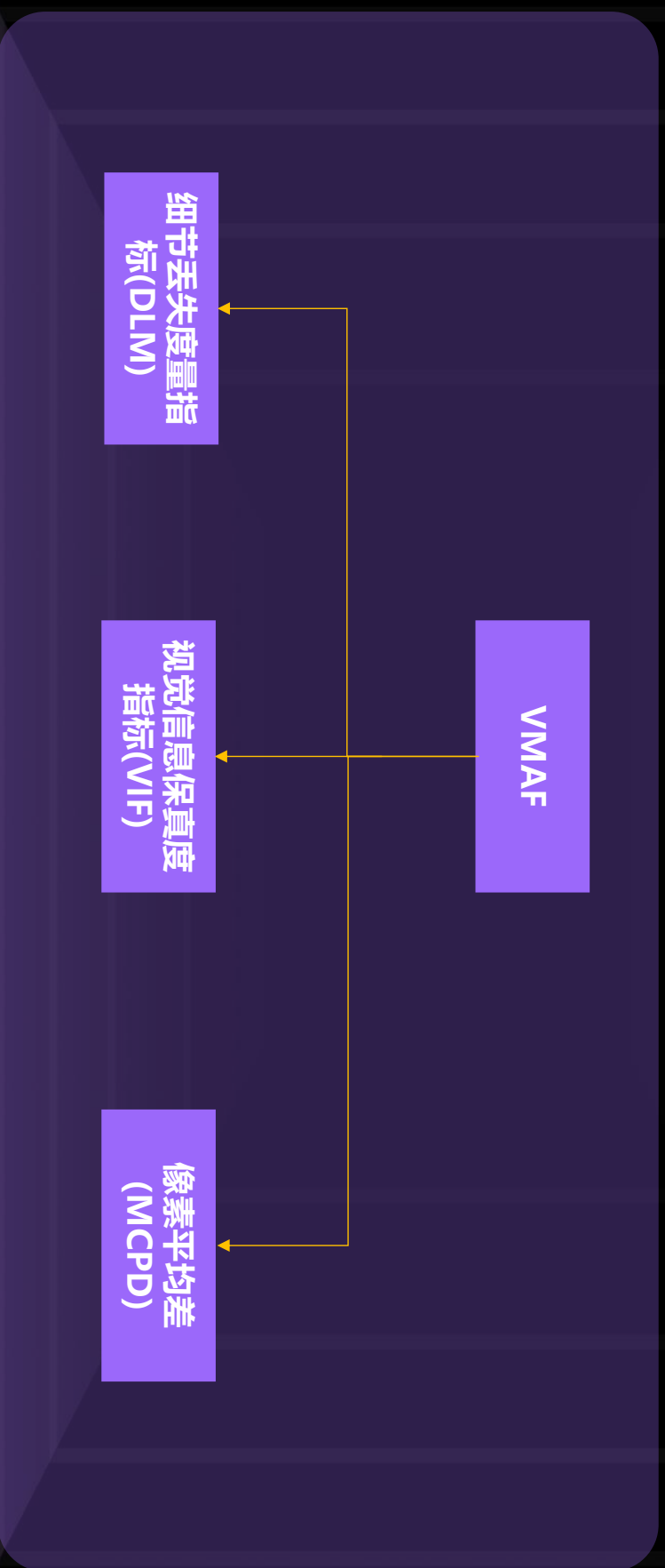
$$PSNR = 10 * \log_{10} \left(\frac{MAX_I^2}{\sqrt{MSE}} \right)$$

SSIM

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

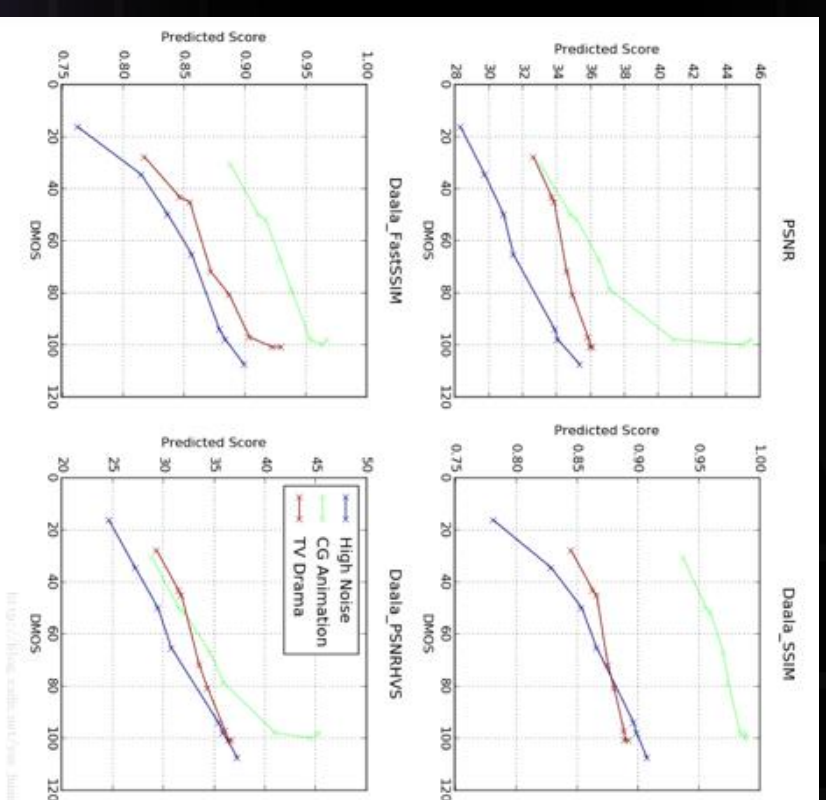
音视频客观评价标准

VMAF



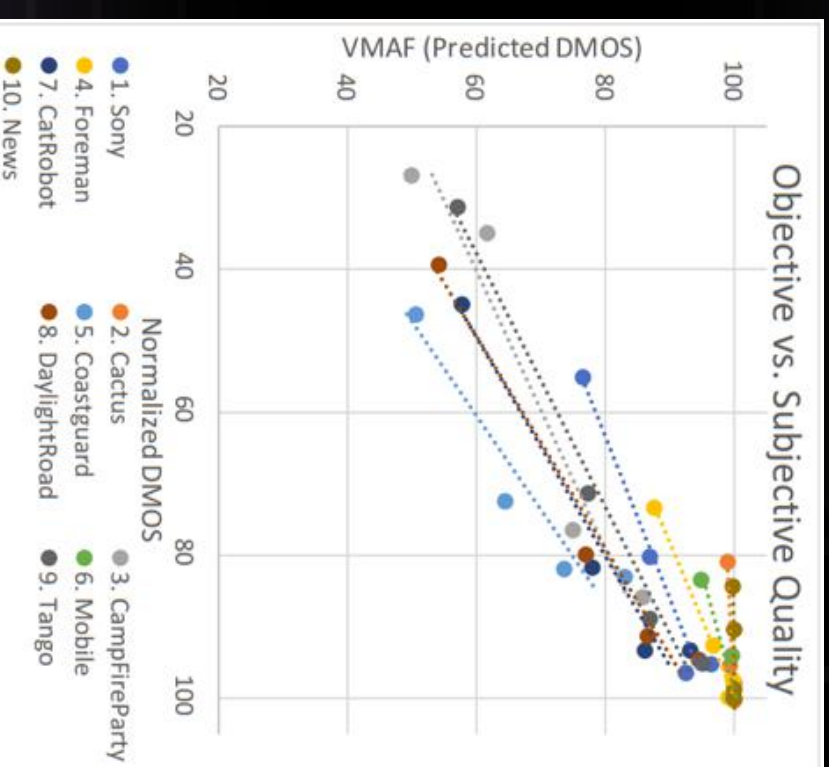
音视频客观评价标准

PSNR/SSIM 分数和 DMOS 关系图：



音视频客观评价标准

VMAF 分数和 DMOS 关系图:

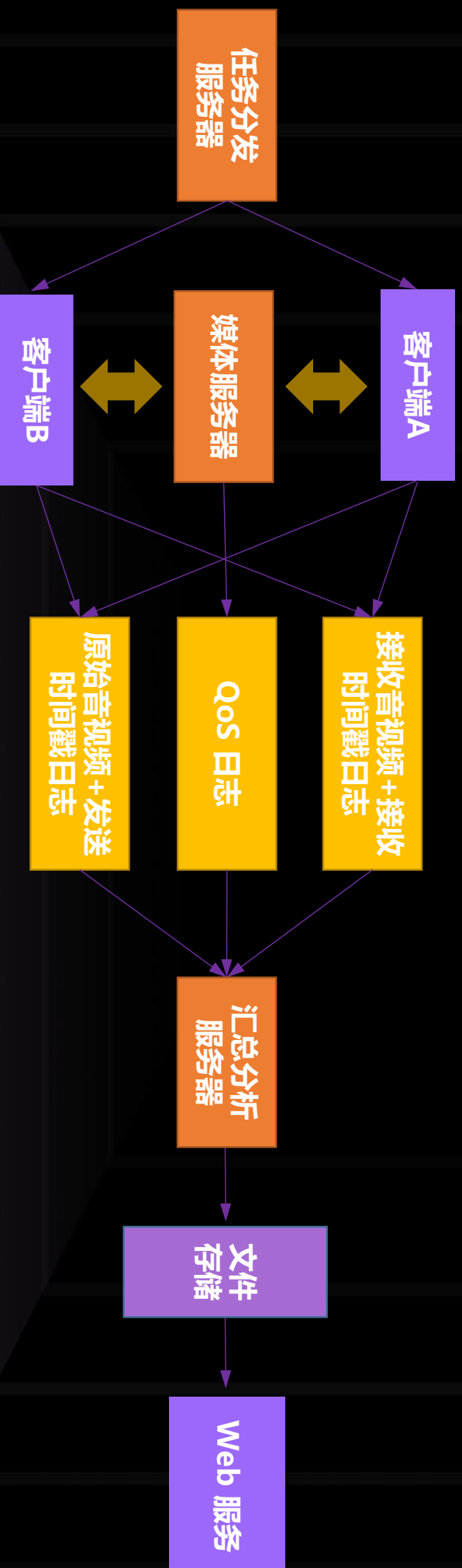


目录

- 1 为什么需要质量评估
- 2 音视频质量评估
- 3 音视频客观评价标准
- 4 基于客观指标的自动化测试**
- 5 音视频质量评价的挑战与未来

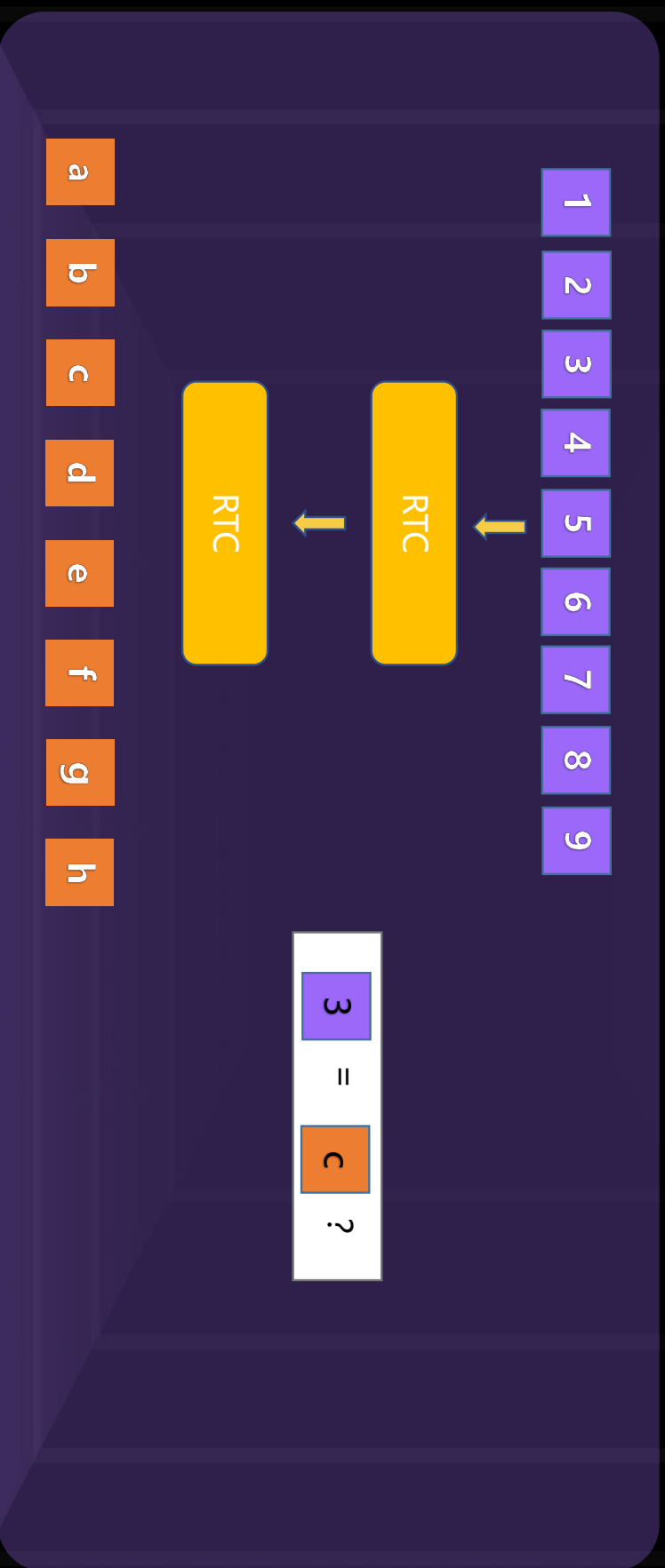
基于客观指标的自动化测试框架

自动化测试总框架：



基于客观指标的自动化测试框架

丢帧的影响：

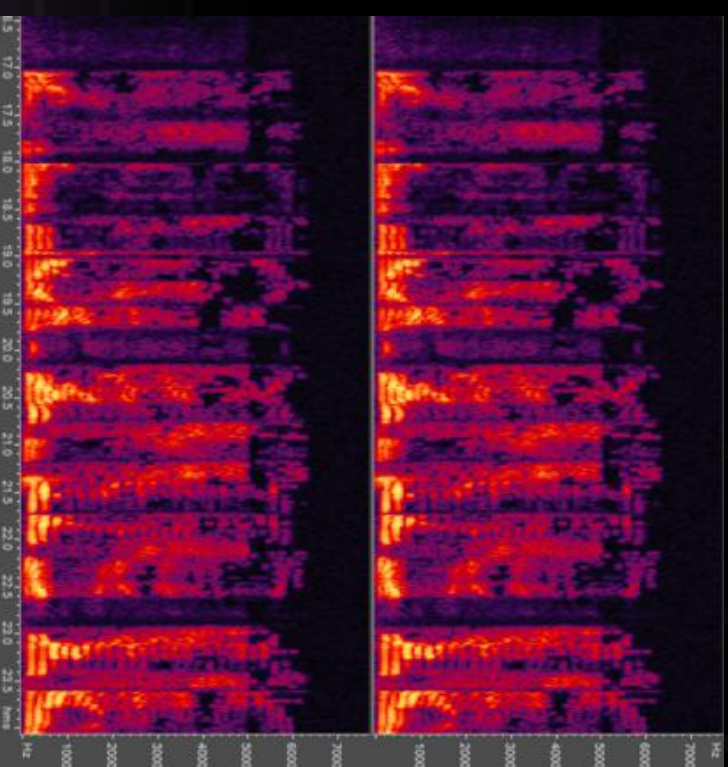
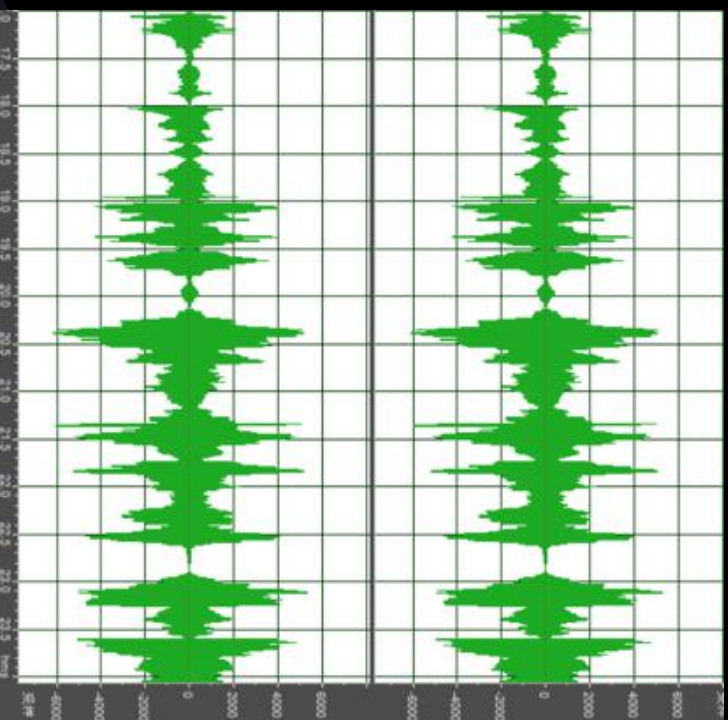


基于客观指标的自动化测试框架



基于客观指标的自动化测试框架

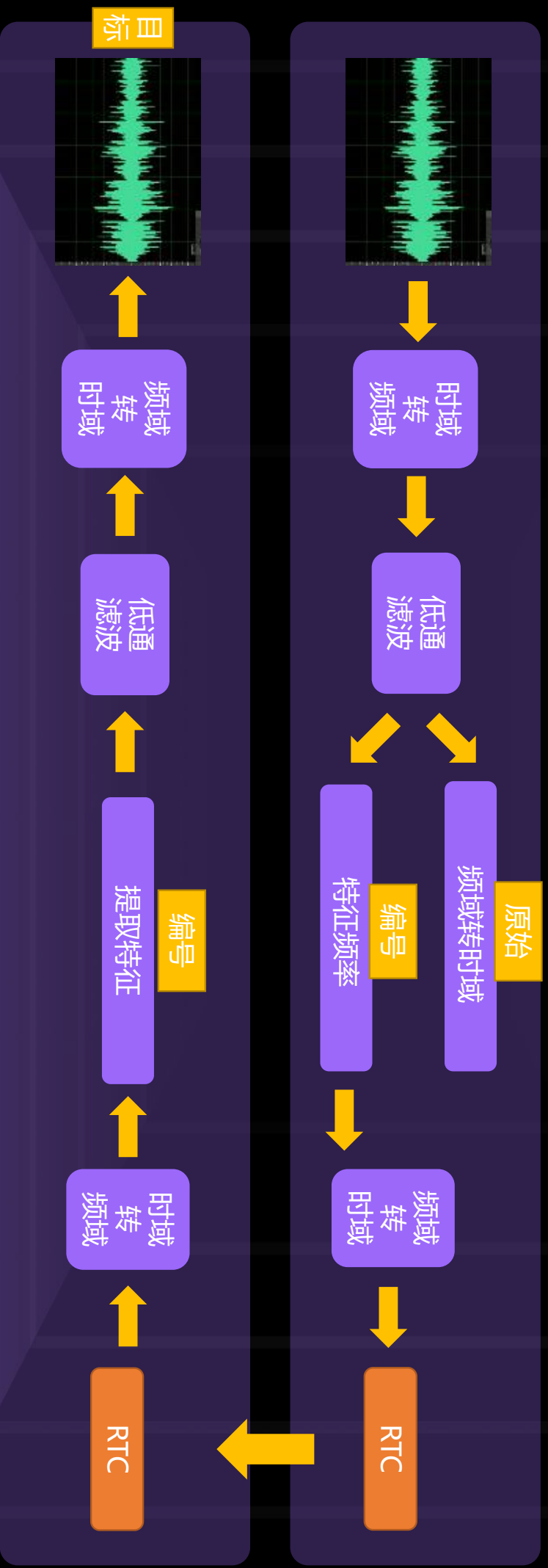
音频语音频率特征：



语音主要集中在4K频率以下

基于客观指标的自动化测试框架

音视频对齐方案：



基于客观指标的自动化测试框架

多端时钟对齐问题：



1v1通话500ms延时
就会影响交互体
验



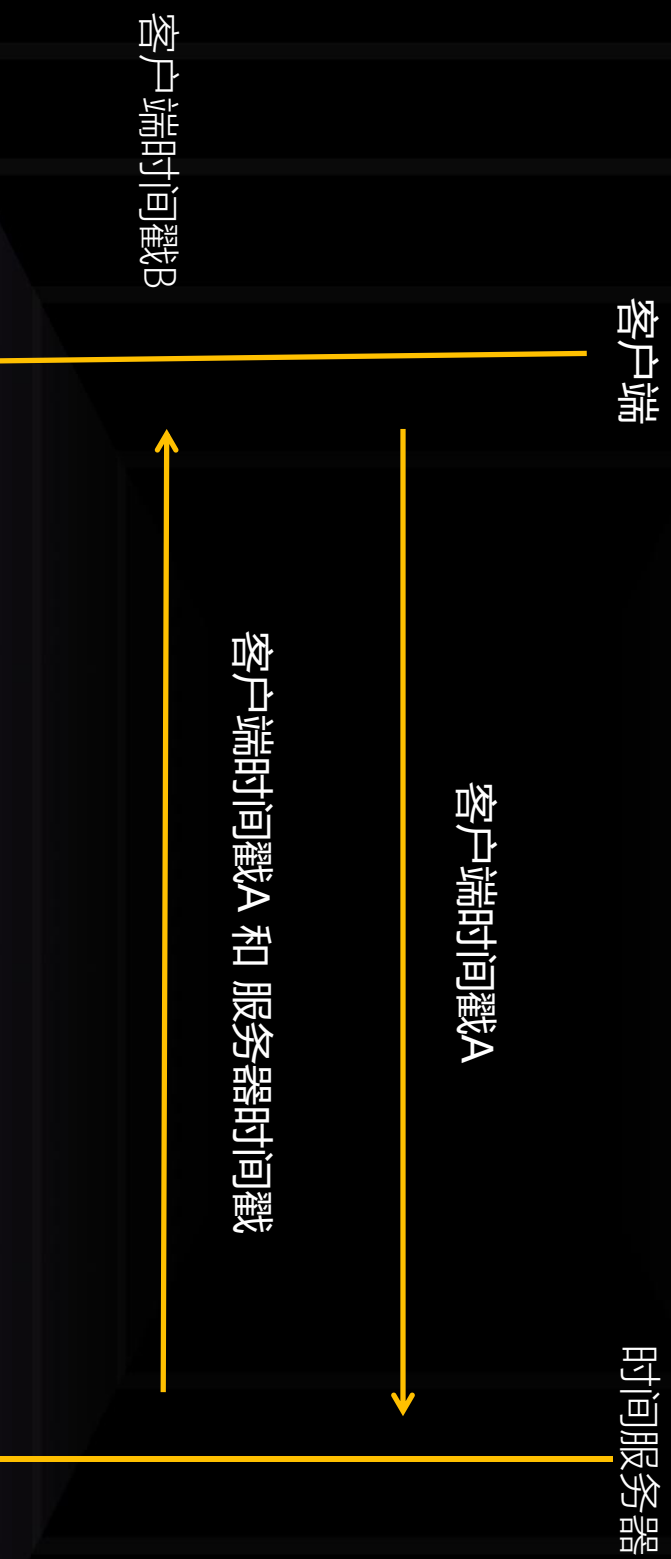
不同设备的时钟差
异在秒级别



需要同步多端的时
钟误差100ms内

基于客观指标的自动化测试框架

时钟对齐方案:



$$\text{修正时间} = \text{服务器时间戳} + (\text{客户端时间戳B} - \text{客户端时间戳A}) / 2$$

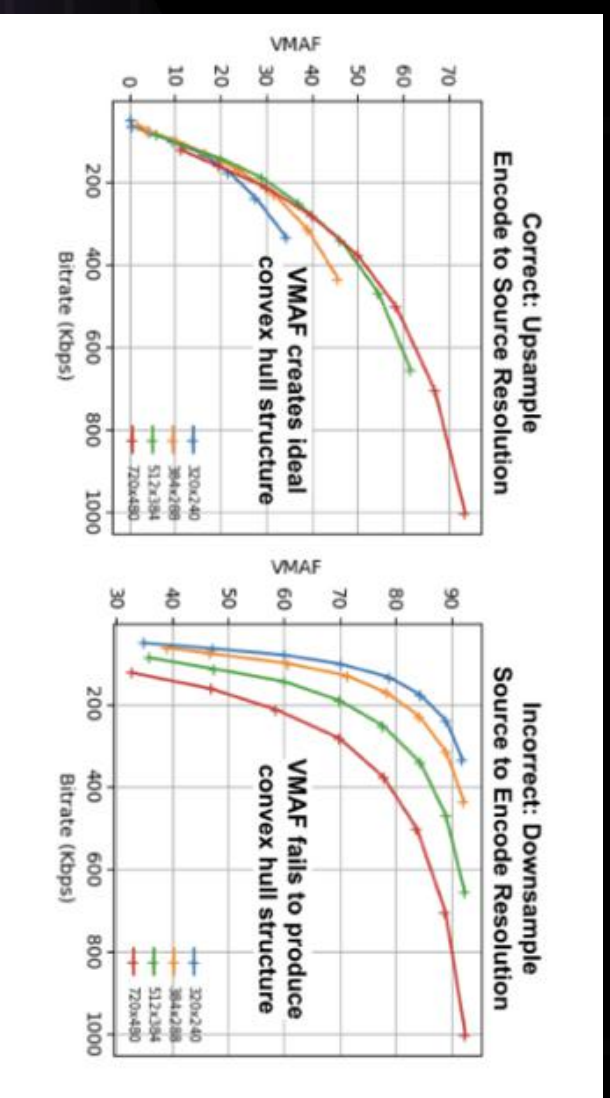
基于客观指标的自动化测试框架

分辨率变化问题:

- 传输过程中，由于网络状况变化以及端侧性能等问题，会导致端上的 SDK 发送低于原始视频分辨率的视频。
- VMAF 要求对比的接收视频和原始视频分辨率必须一致，这就需要把接收分辨率上采样到原始分辨率，或者把原始分辨率下采样到接收分辨率。
- 计算 VMAF 的正确方法是将接收分辨率上采样到原始分辨率。

基于客观指标的自动化测试框架

如下图，当上采样到原分辨率计算 VMAF 时，可以轻松识别出不同分辨率曲线之间的交点。相反，如果下采样到编码分辨率计算 VMAF，则低分辨率编码将产生不匹配的高分，并且曲线之间不会出现相交。



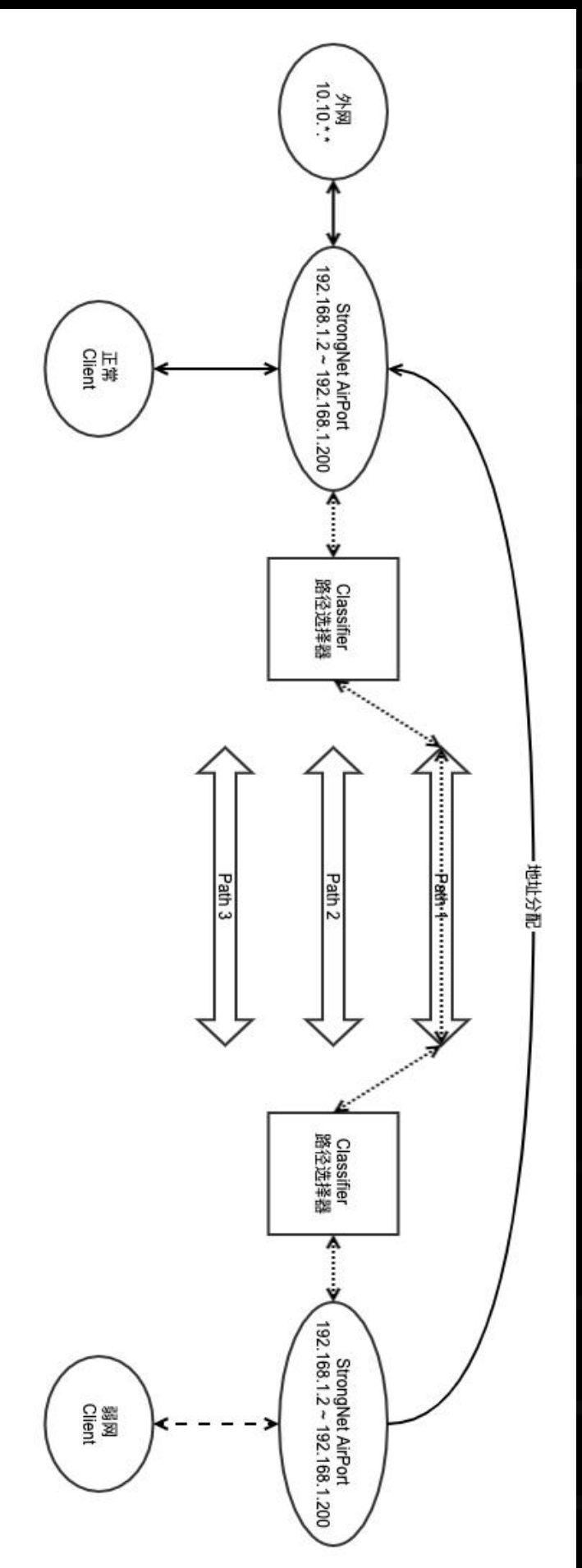
基于客观指标的自动化测试框架

弱网模拟：

- 可通过模拟不同的网络环境，一方面验证 SDK 优化的各项性能指标是否合格，另一方面验证弱网环境下的音视频质量。
- 实践中，客户端使用 HoloWan 弱网模拟环境；媒体服务器采用 Docker 部署，使用 tc/netem 命令直接模拟。

基于客观指标的自动化测试框架

Holowan 弱网环境网络拓扑：



基于客观指标的自动化测试框架

分析方式:

发送方

帧号	时间戳	图像
1	1634401000	二进制
2	1634401100	二进制
3	1634401200	二进制
4	1634401300	二进制
5	1634401400	二进制
6	1634401500	二进制

+

接收方

帧号	时间戳	图像
1	1634401000	二进制
2	1634401100	二进制
3	-	-
4	-	-
5	1634401400	二进制
6	1634401500	二进制

=

分析结果

指标
丢帧情况
帧率,流畅度
全参考图像质量
延迟
抖动

基于客观指标的自动化测试框架

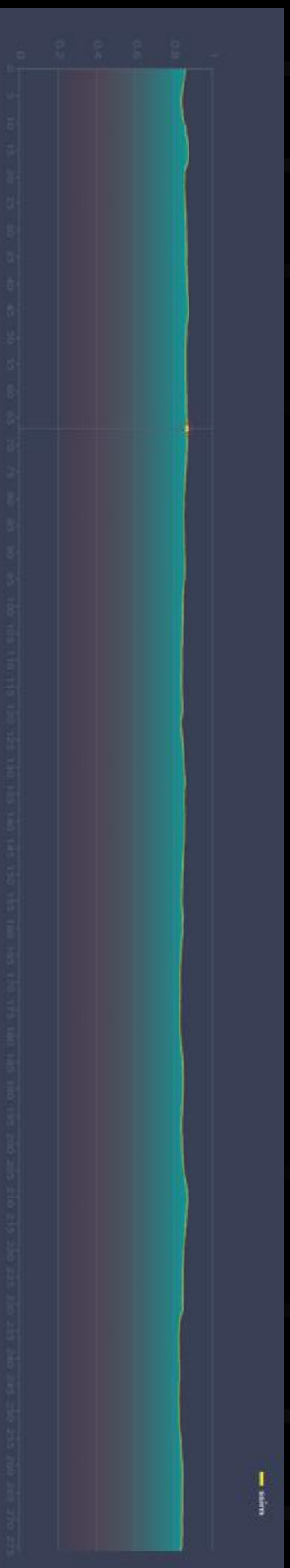
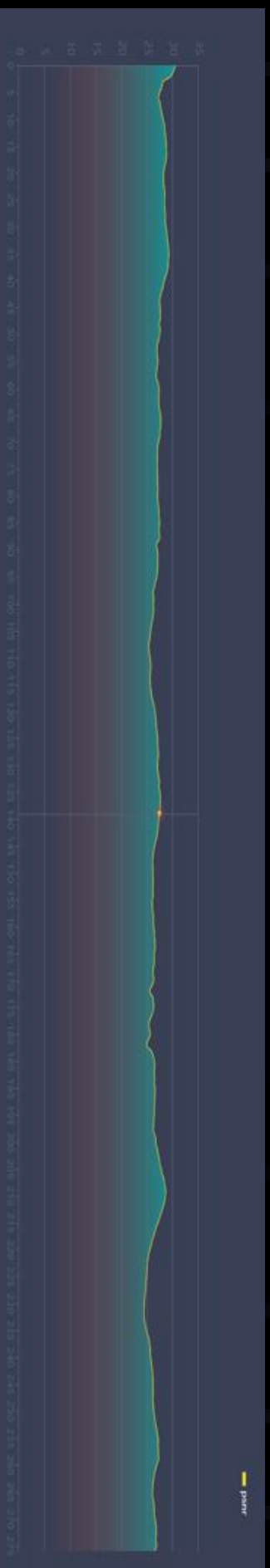
结果展示:

自动化测试结果分析完成后，可以通过 Web 页面访问任务列表，并在任务列表中按照客户端唯一标识查看对应的分析结果图。

客观指标评估结果		客观指标评估结果	
Room ID	用户 ID	查询	
房间 ID	用户 ID	创建时间	操作
9c1056c7-7adb-40f8-bc8f-0637133058ac	bce945f658aa40689b7637906829ee34	2022-01-10 09:29:55	查看结果
9c1056c7-7adb-40f8-bc8f-0637133058ac	e7f5a3fecd44362a1e0a94a0d240a2	2022-01-10 09:22:29	查看结果
9b0f6102-5e4e-451b-9d75-6c4e87b2a134	8e5f65c4d0f14b4da85b5b10074c6c88	2022-01-10 06:44:09	查看结果
9b0f6102-5e4e-451b-9d75-6c4e87b2a134	f59d5f0e1e34b5dbaec054f0a75f4d	2022-01-10 06:40:28	查看结果
36cd47a1e6-4f5f-a197-ac05db96640b	7da3d926-5333-4719-886b-aad4123c369a	2021-09-23 06:55:36	查看结果
b53787c9-d3b0-45b4-93f6-e4ae28ea5e0	e3839090-4021-4d95-a0c-a0d571ddc69	2021-09-23 05:51:58	查看结果
5e5f797b-aba6-47ea-9816-cb14e40da7bd	05ba7967-fa16-4b19-a8e4-7d2789a9a7f4	2021-09-23 05:13:06	查看结果
9d3917d6-856a-447a-93fb-3e5e8bdeb3be	4967931c-9446-4d42-b89c-aed182e6808a	2021-09-23 01:32:31	查看结果
832b1631-2077-4c14-8b84-276fb25b5a8e	403c23a1-7dcd-4532-8c99-e236750d3a67	2021-09-22 18:34:41	查看结果

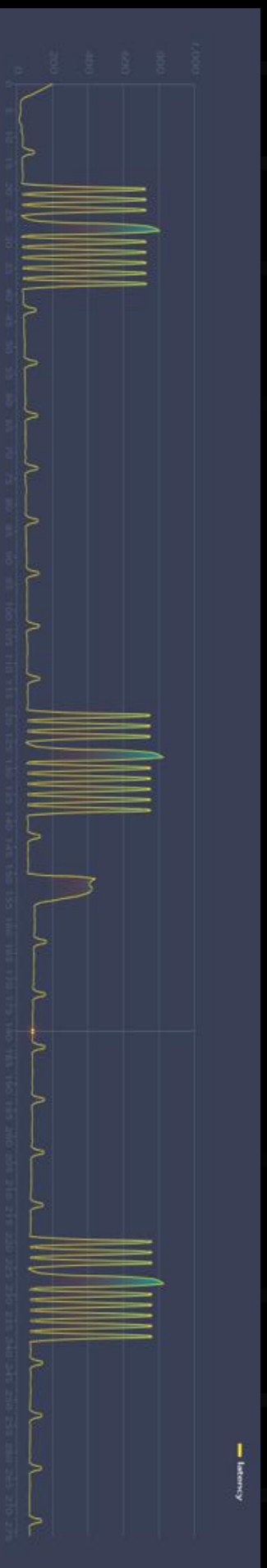
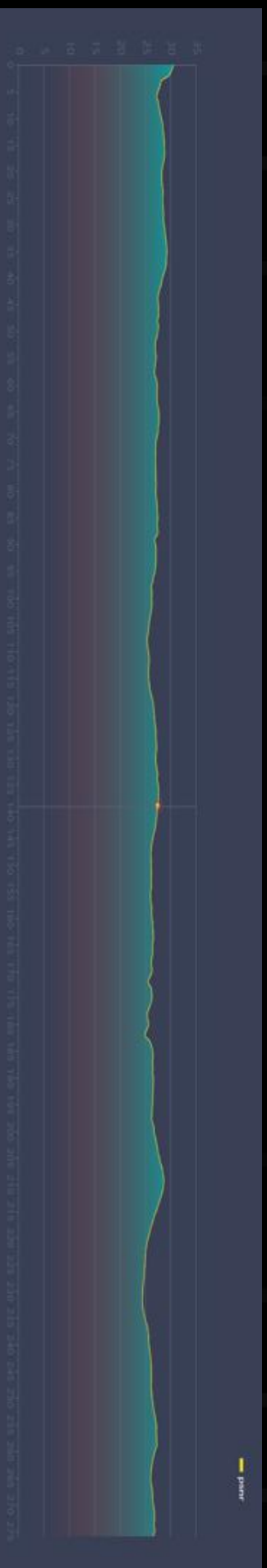
基于客观指标的自动化测试框架

结果指标:



基于客观指标的自动化测试框架

结果指标:



基于客观指标的自动化测试框架

资源工具和清单:

- WebRTC: www.webrtc.org
- PESQ: <https://www.itu.int/rec/T-REC-P.862>
- POLQA: <http://www.polqa.info/products.html>
- VisQOL: <https://github.com/google/visqol>
- FFmpeg: <https://ffmpeg.org>
- SoX: <http://sox.sourceforge.net>
- OpenCV: <https://opencv.org>

目录

- 1 为什么需要质量评估
- 2 音视频质量评估
- 3 音视频客观评价标准
- 4 基于客观指标的自动化测试
- 5 音视频质量评价的挑战与未来

音视频质量评价的挑战与未来

面临的挑战

音视频质量评价在业界经历了十几年的发展，但它的应用依然不多。原因主要有：

- 算力成本高
- 方法本身应用场景限制
- 核心问题不在于评价
- 投入产出比低

音视频质量评价的挑战与未来

结合当前的现状，个人认为未来发展的一些可能的方向：

- 在 PSNR 和 SSIM 基础上构建新指标
- 视听质量评估
- 非参考质量评估

感谢聆听!



融云官方公众号



系列课程讨论群