

直面音视频质量评估之痛

——走进腾讯音视频质量体系

腾讯/音视频实验室/罗必达



腾讯视频实验室：让世界在你身边



QQ音视频日均通话 **12亿分钟**



基于腾讯云，服务 **130+家** 客户

高质量的音视频通话技术离不开相应的**评估与测试体系**

音视频质量评估是干啥的？

音视频技术开发和测试都必须具备的**基本技能**

评估的对象包括：

- 传输与流控的质量
- 编解码的质量
- 数字信号前后处理的质量
- 设备兼容的质量
- 上线后音视频质量的监控

三大痛点

主观性

主观测试与客观测试相结合，合理利用客观测试

不确定性

真实环境验证最终质量，通过实验室消除不确定性

矛盾与平衡

综合性指标，结合其他维度建立全维度数据分析及监控能力

音视频测试面临的挑战

PASS or **FAIL** ?

音视频测试面临的挑战

主观性

一百个人心中有一百个哈姆雷特



音视频测试面临的挑战

矛盾与平衡

鱼与熊掌不可兼得

可用资源有限的情况下：

清晰度

流畅度

噪点

亮度

网络抗性

通话延时

超清

耗电

带宽探测精度

带宽探测速度



音视频测试面临的挑战

不确定性

不可能两次踏进同一条河流

信号的不确定性

环境的不确定性



主观性

主观测试与客观测试相结合，**合理利用客观测试**

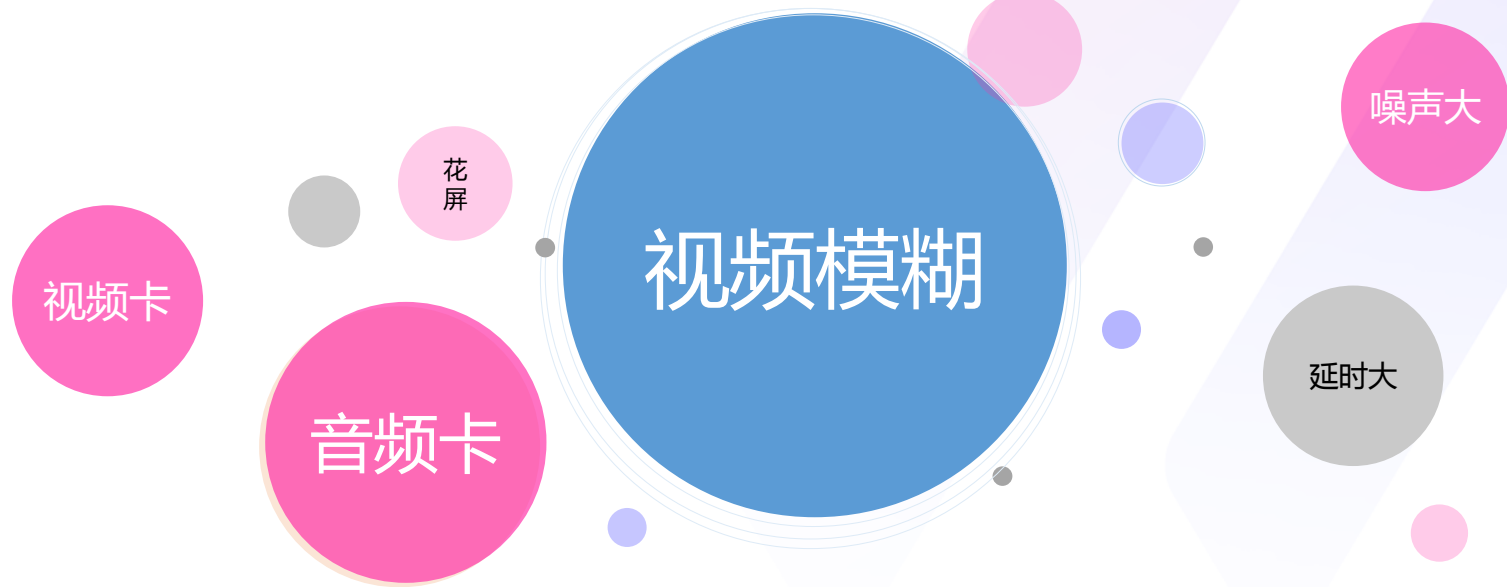
不确定性

真实环境验证最终质量，通过实验室消除不确定性

矛盾与平衡

综合性指标，结合其他维度建立全维度数据分析及监控能力

通话质量差，你想到了哪些现象？



基本指标

类别	关注指标
视频	清晰度、帧率（含卡顿）、延时、色彩准确性、分辨率
音频	音质、回声抵消、噪声抑制、增益控制、延时、频响
综合	音画同步、性能消耗

主观测试

- 反映真实感受，衡量音视频质量的最终手段
- 主观个体差异大
- 时间人力成本高，重复性低，数据缺乏参考性

客观测试

- 数据量化，参考性高，快速响应

案例：

- 视频清晰度测试
- 音频端到端延时测试

案例1：视频清晰度测试



视频清晰度 \neq 视频分辨率

案例1：视频清晰度测试



如何量化？

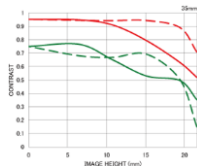
需要**客观测试**进行**辅助**

案例1：视频清晰度测试

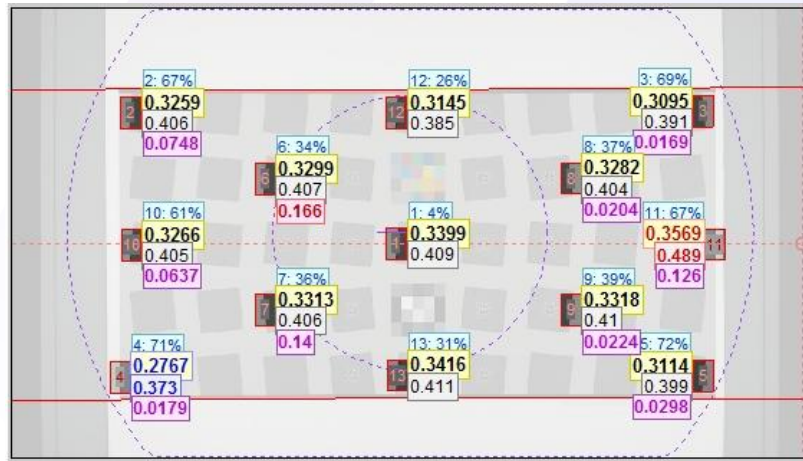
任意视频内容



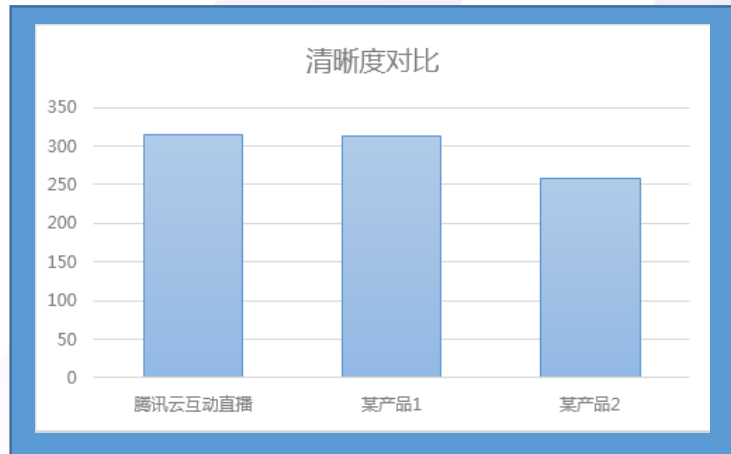
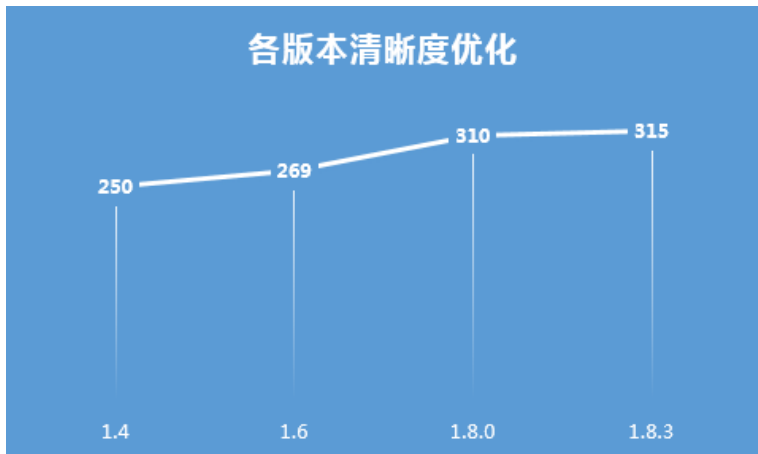
固定视频内容



MTF50



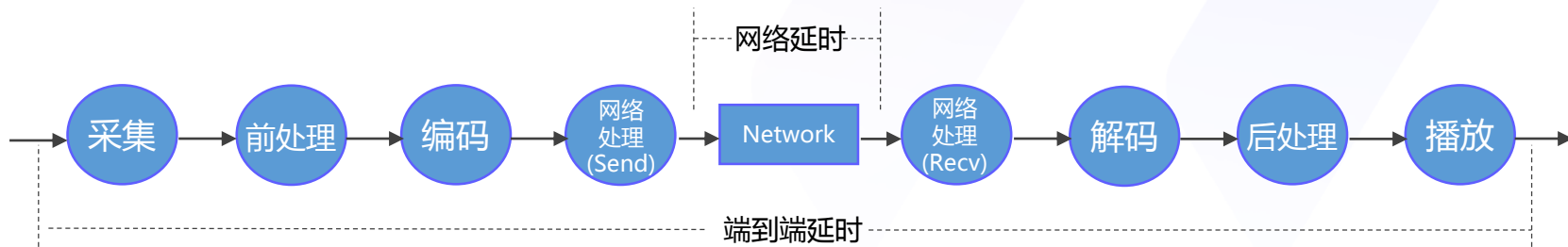
案例1：视频清晰度测试



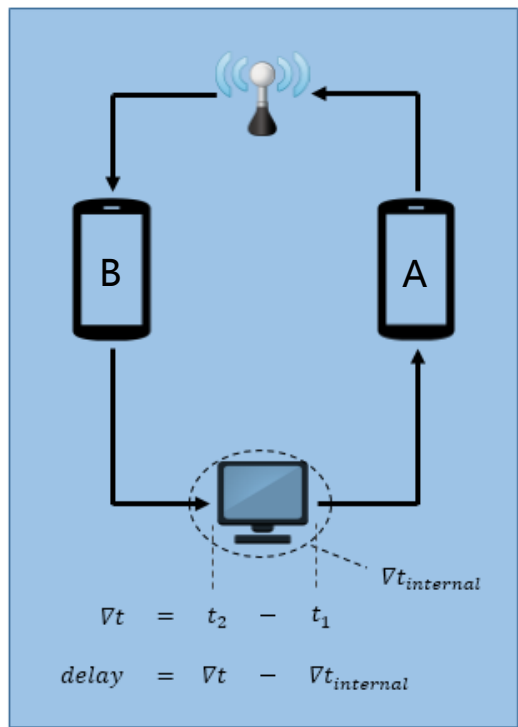
主观测试工作量：**1天** → 客观测试工作量：**0.5小时**

案例2：音频端到端延时测试

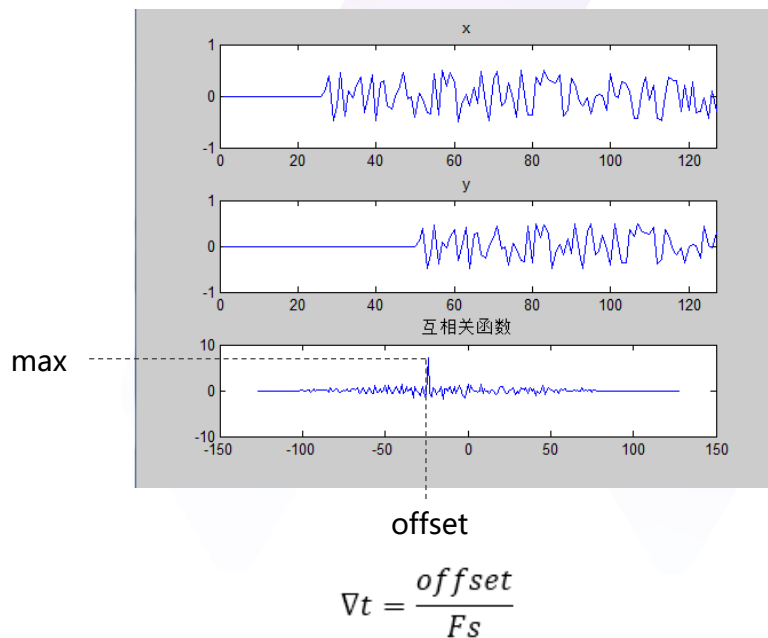
有些测试主观无法完成，**必须借助客观测试的手段**



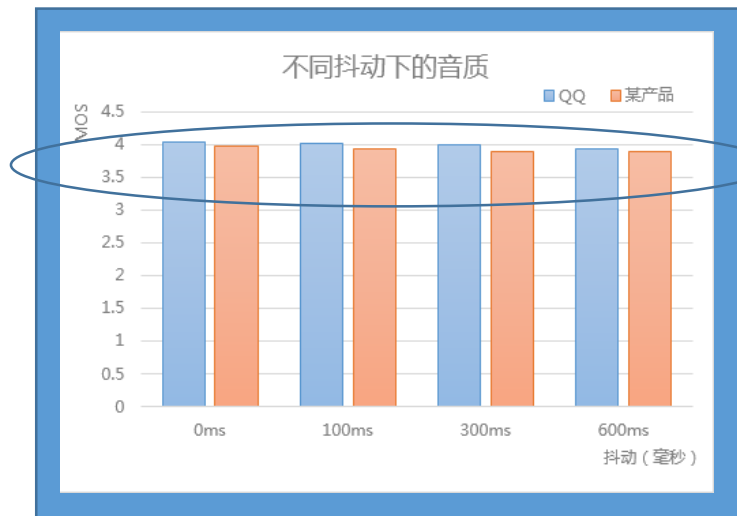
案例2：音频端到端延时测试



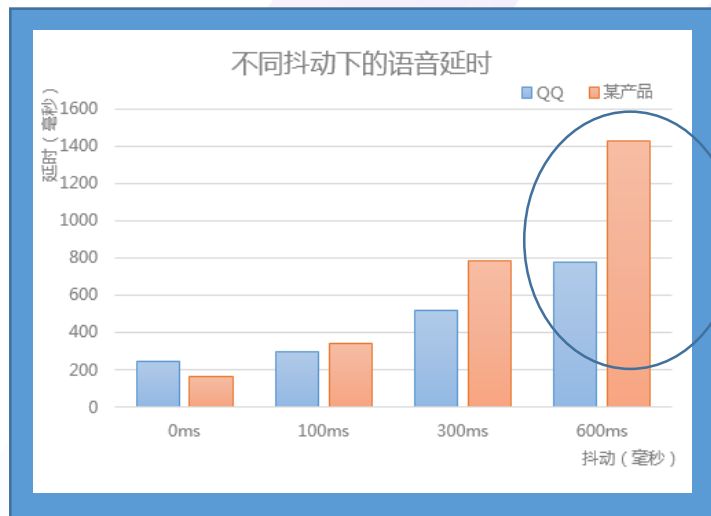
互相关函数： $f(t) = \sum x(n)y(n+t)$



案例2：音频端到端延时测试



不同抖动下音质平稳



用延时换取了抖动抗性

更需要关注的是：在不同场景和环境下，各指标的情况是怎样的

Q: 环境的不确定性！

主观性

主观测试与客观测试相结合，合理利用客观测试

不确定性

真实环境验证最终质量，**通过实验室消除不确定性**

矛盾与平衡

找到综合性指标，结合其他维度建立全维度数据分析能力

为什么我们需要一个专业的实验室

每次测试结果都不一样？

Bug无法重现？

测试结果没有指导意义？

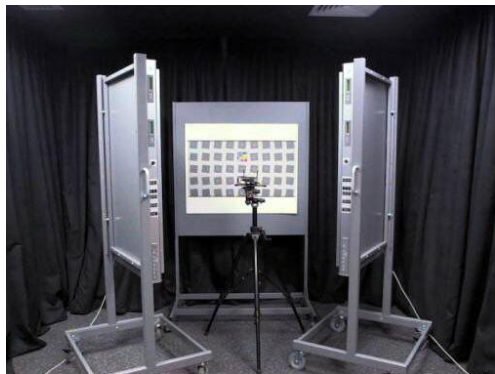
工欲善其事，必先利其器！

为什么我们需要一个专业的实验室



目的：使测试数据更具指导意义

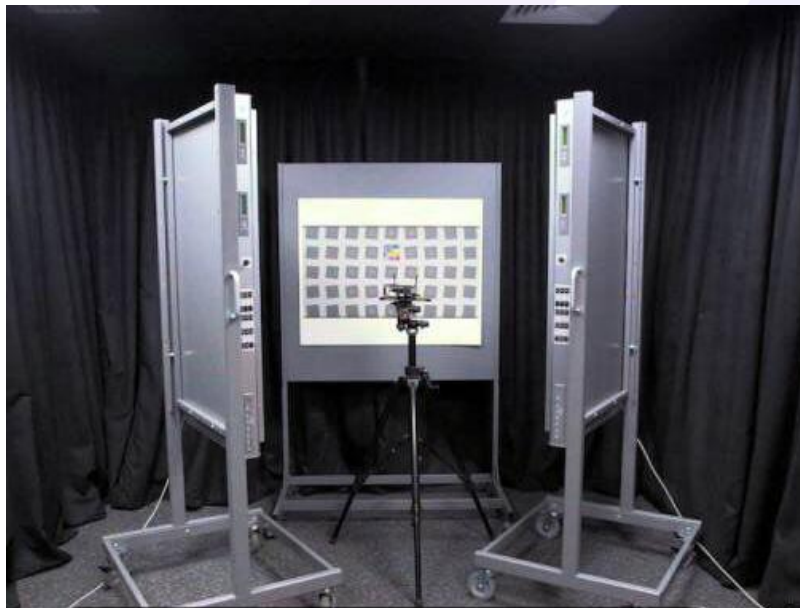
音视频测试实验室全貌



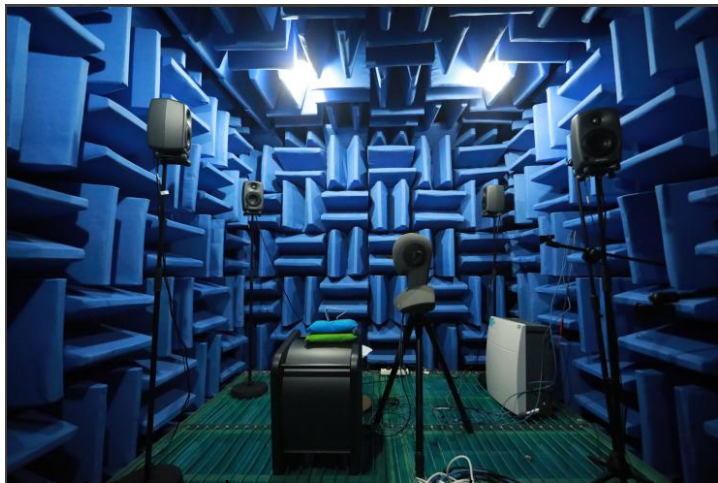
一个基本的实验室需要具备哪些条件

精确控制灯光

- 照度、色温



一个基本的实验室需要具备哪些条件



精确控制声音输入

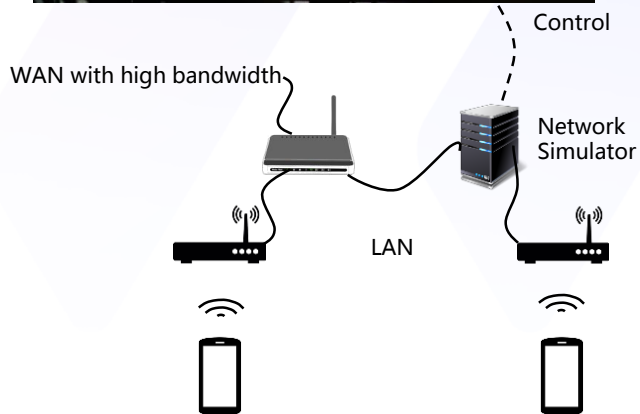
- 音源、音量、噪声



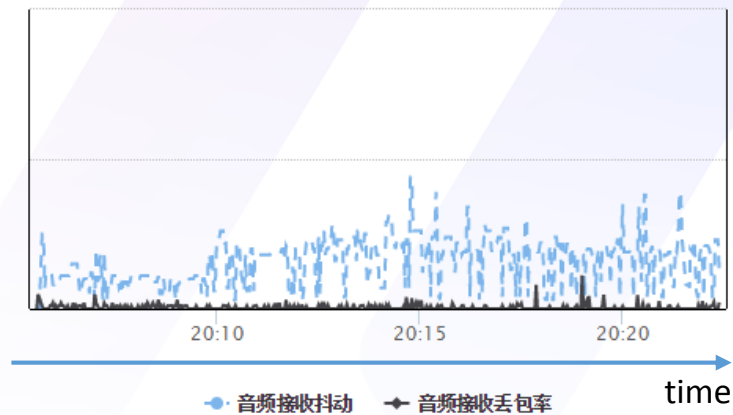
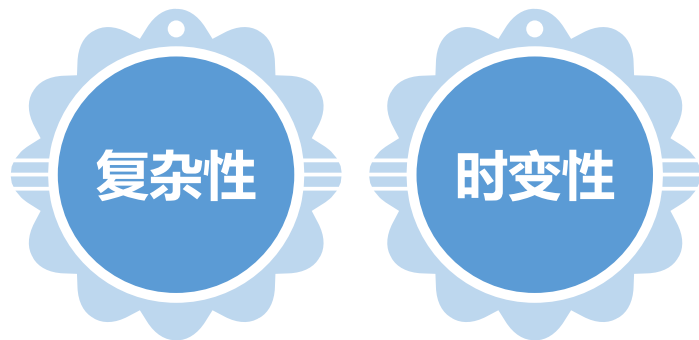
一个基本的实验室需要具备哪些条件

精确控制网络损伤

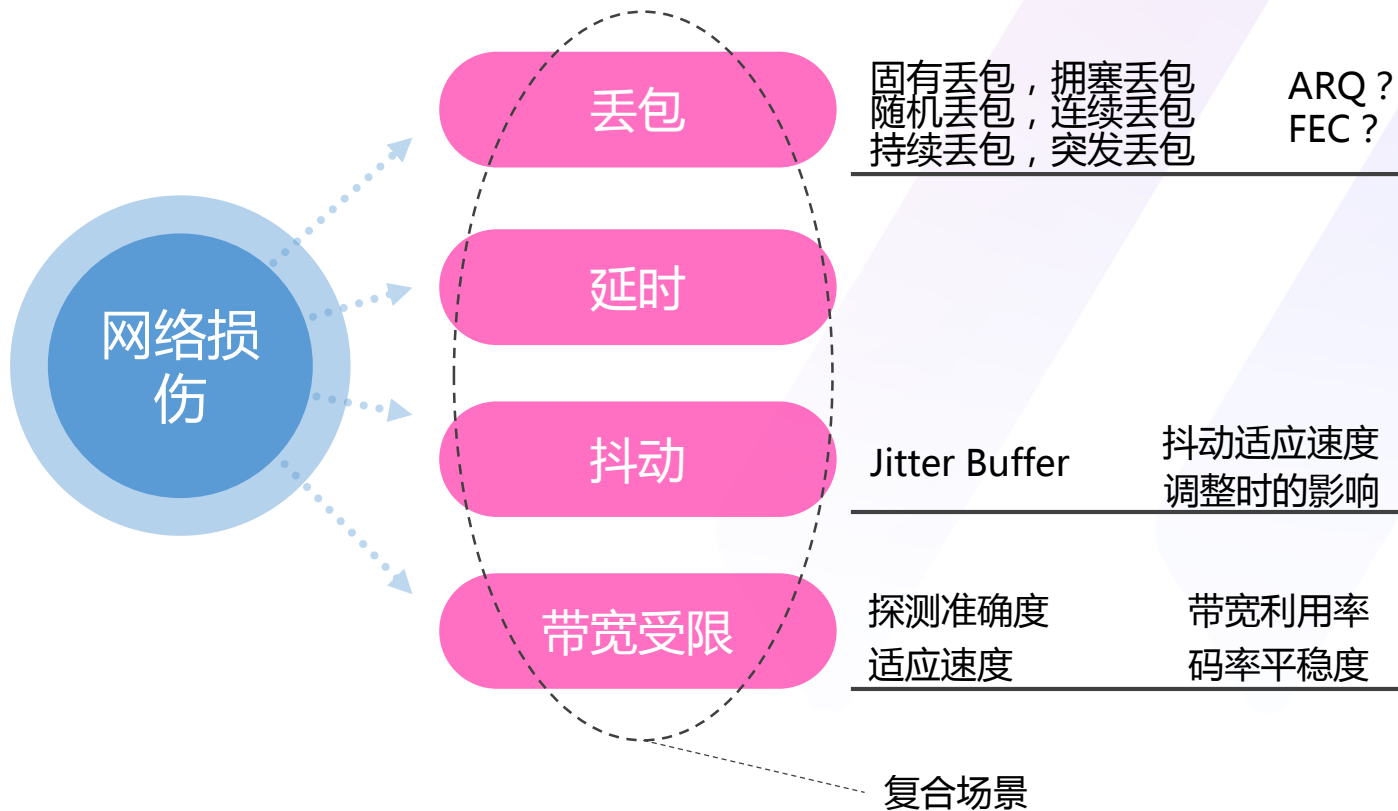
- 无损、带宽、丢包、抖动、延时



网络抗性

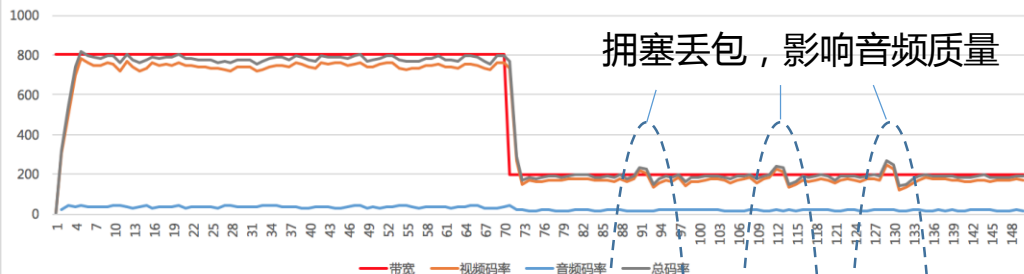


常见网络损伤



案例3：音视频联合流控

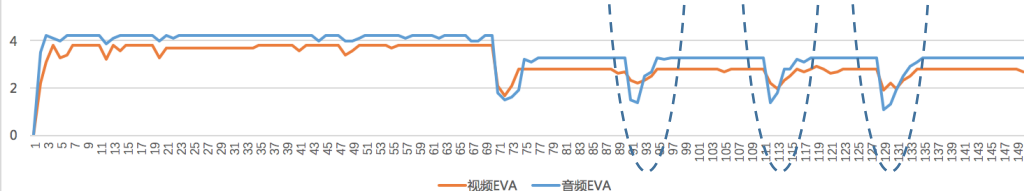
带宽受限下音视频码率变化情况



丢包率



音视频质量



应对方案：优先保证音频质量

- 优先降低视频码率
- 使用音频冗余试探带宽
- 不使用丢包率作为带宽探测判断依据

实验室测试就够了吗？

Q：与真实环境有差异？

实验室测试：明确指出缺陷所在，指导优化方向

真实环境测试：最终结果验收

Q：如何对现网做音视频质量监控呢？

主观性

主观测试与客观测试相结合，合理利用客观测试

不确定性

真实环境验证最终质量，通过实验室消除不确定性

矛盾与平衡

综合性指标，结合其他维度建立**全维度**数据分析及监控能力

通话质量？用数据说话

老板

能从大盘上看看最近优化对用户有什么影响吗？

老板你看，平均丢包率
减少了3个百分点

员工

老板

但用户体验有提升吗？

帧率分布也上去了

员工

老板

但用户体验有提升吗？

这...



员工

开发

这次经过我的优化，视频
流畅多了吧 😊

扯淡！画面模糊了这么多！

测试

开发

可用码率就这么多，你
不能要求这么多啊...

但你拿清晰度换流畅度，
算什么优化啊！

测试

老板

谁能告诉我优化究竟有没有
效果？！

开发

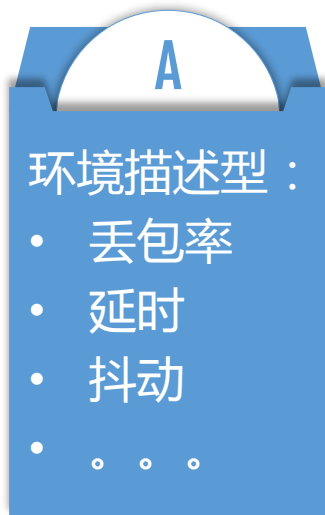
这...

这...

测试

原因：**缺少一个综合评价数据**，以表征音视频质量对于用户的感受

通话质量？用数据说话



EVA：综合各维度的结果型无参考评估模型

什么是无参考评估

A

全参考评估

- **参考原始信号**
- 准确性更高
- 不适用于原始信号增强类需求
- PSNR , SSIM , PESQ , POLQA

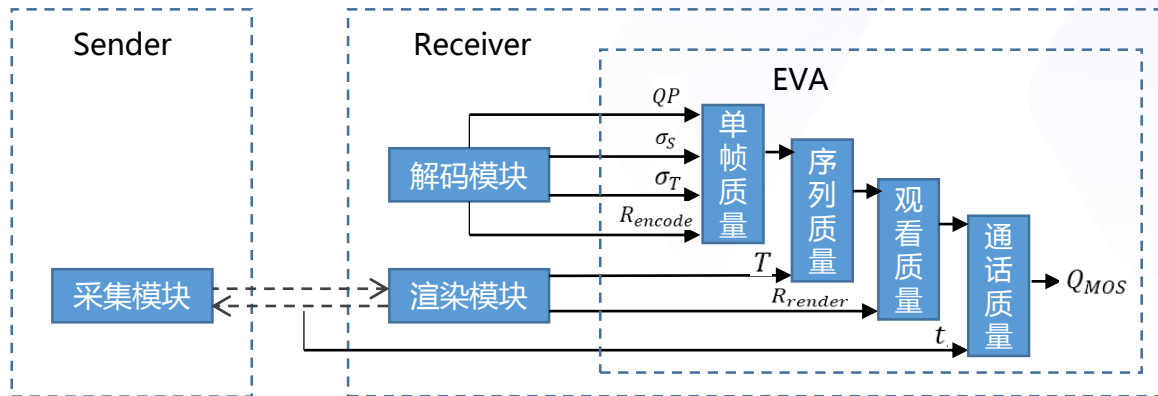
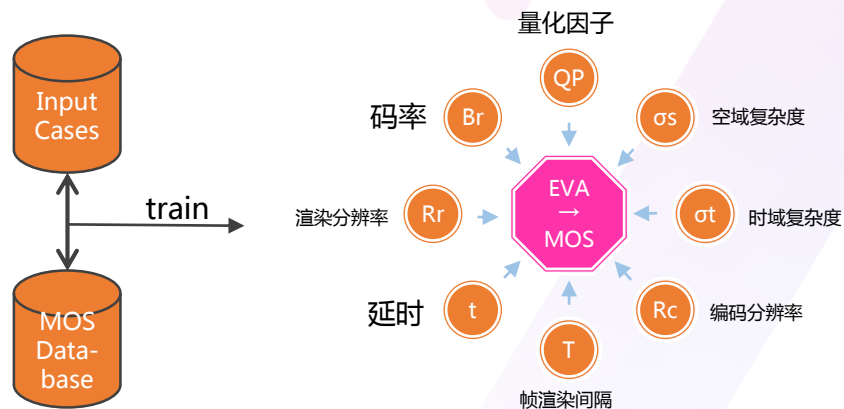
B

无参考评估

- **不参考原始信号**
- 准确性相对较低
- 更适合线上评估
- G.1070

线上质量评估，无法获取原始信号

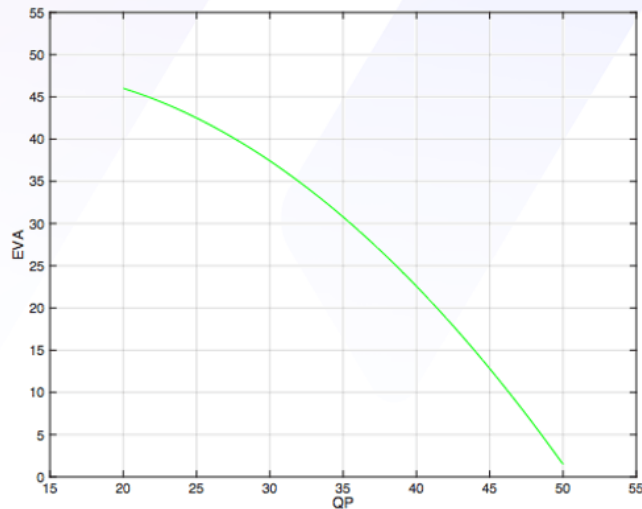
无参考评估体系EVA-以视频为例



无参考评估体系EVA-QP



实验发现，QP与单帧质量**强相关**



无参考评估体系EVA-内容参数

内容参数：空域复杂度 σ_S & 时域复杂度 σ_T

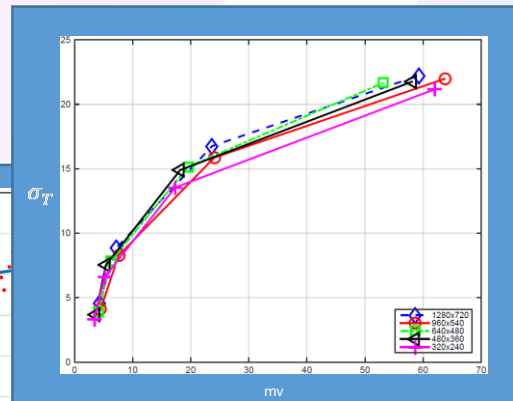
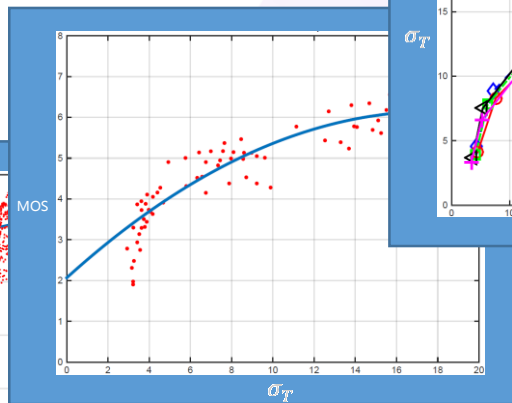
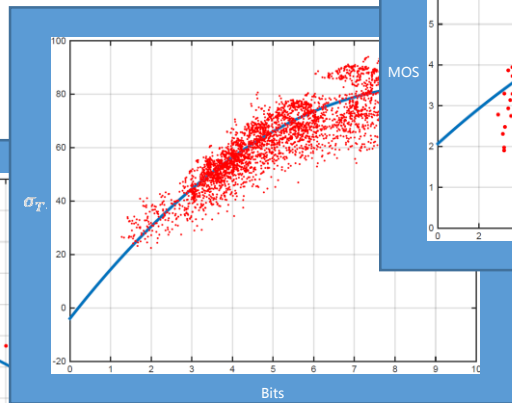
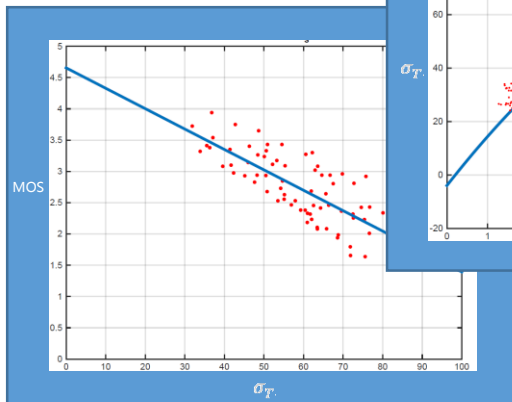
Frame :



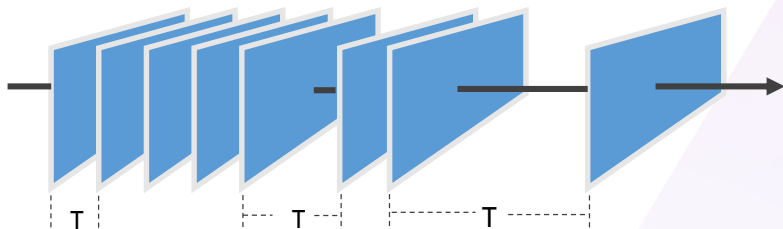
σ_S :



无参考评估体系EVA-内容参数



无参考评估体系EVA-流畅度与延时



帧渲染间隔T：

- 帧率
- 帧率抖动
- 卡顿

延时

1

延时对EVA的影响视应用场景而定

2

延时大小对交互的影响程度

3

单向延时 or 双向延时

综合性指标

开发

这次经过我的优化，视频流畅多了吧 😊

扯淡！画面模糊了这么多！

测试

开发

可用码率就这么多，你不能要求这么多啊...

但你拿清晰度换流畅度，算什么优化啊！

测试

老板

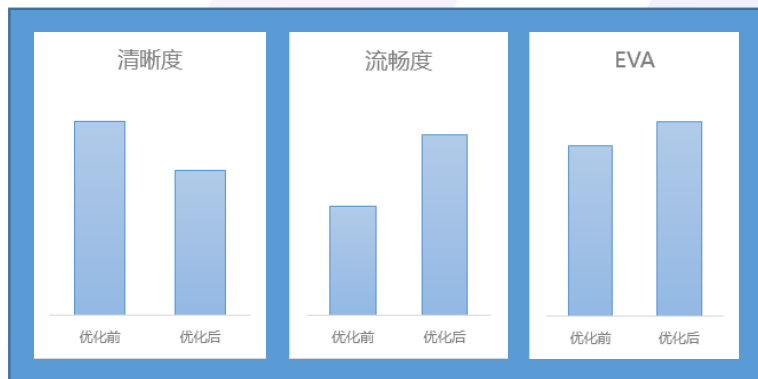
谁能告诉我优化究竟有没有效果？！

开发

这...

这...

测试



开发

那我们就看EVA分数吧！

测试

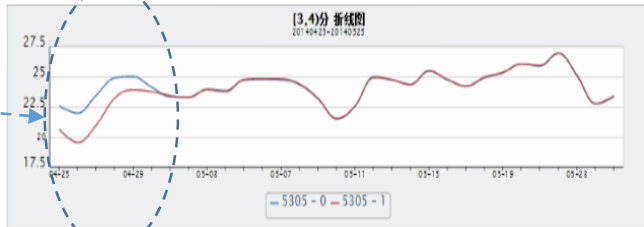
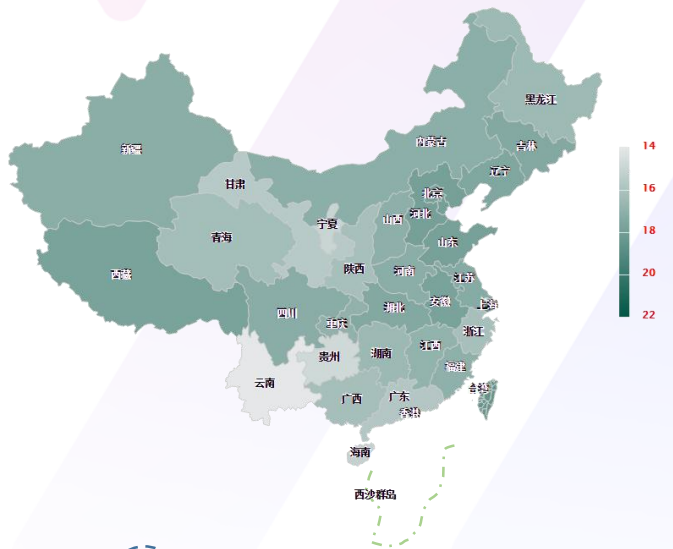
EVA的应用-全局视角

EVA优缺点：

- 能够较高拟合度地反映用户体验
- 反映结果，无法得知造成该结果的原因
- 分析时需要配合环境描述数据及单维度数据

应用：

- 分时、分区域观测全网音视频通话质量
- 评估后台接口机部署质量
- 版本上线监控
- A/B Test



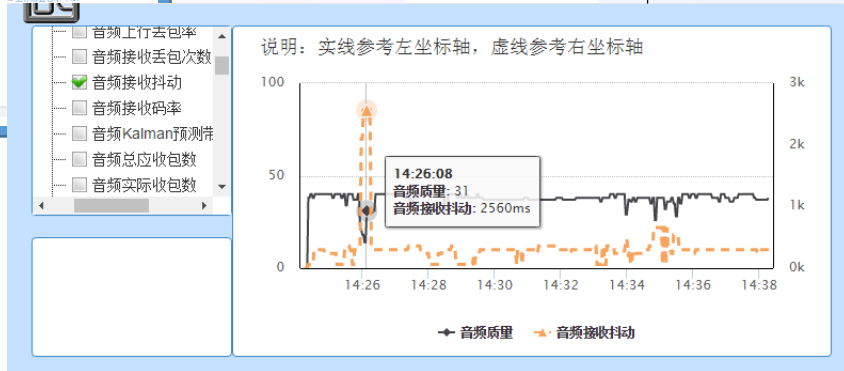
个例分析



个例分析：具体问题具体分析解决

线上实时单例数据监控

线下单例数据分析



全局与个例分析相结合，建立全维度数据分析与监控能力

总结

- **主观性**

- 不能忽视的主观测试
- 当主观测试无法量化，或者测试精度无法满足需求时，需要客观测试进行辅助

- **不确定性**

- 精确控制测试环境，减小音视频测试的不确定性
- 在真实环境中验证最终结果

- **矛盾与平衡**

- 对不同应用场景建立一套综合各指标的评估模型
- 环境描述性数据和单维度数据也很重要
- 建立全局视角和单例视角的数据分析能力

直面音视频质量评估之痛：

主观与客观相结合，
实验室与现网双管齐下，
综合维度与其他维度一起抓

Thanks you!

腾讯音视频实验室公众号：

