



音视频服务系统建设实践

wuzijing@xiaomi.com

定制音视频方案@引擎团队



———— CONTENT ————

01

团队背景

02

实时通话技术

03

推流 / 播放技术

04

视频编辑技术

总体介绍 - 主要技术服务

播放器、编辑器 非实时

1. 边下边播、预加载（播放器）
2. 用户无感快放、慢放（播放器）
3. GPU加速编辑
4. 多轨视频编辑
5. 服务端视频编辑

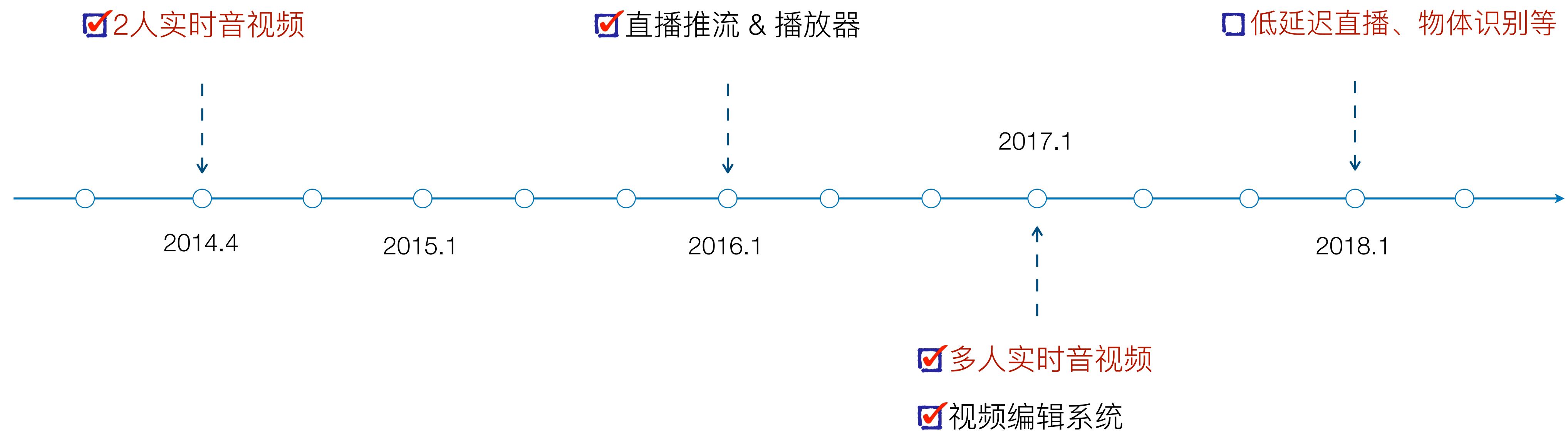
直播推流系统 准实时：delay < 5s

1. 动态贴纸、美颜、大眼瘦脸等特效
2. 混响、混音等声音特效
3. 视频拍摄录制
4. 硬件Codec

实时通话（连麦）系统 实时：delay < 0.5s

1. 双声道**48k**语音实时通话
2. 支持2人ICE穿透通话、多人会议
3. 超大型会议
4. AEC处理

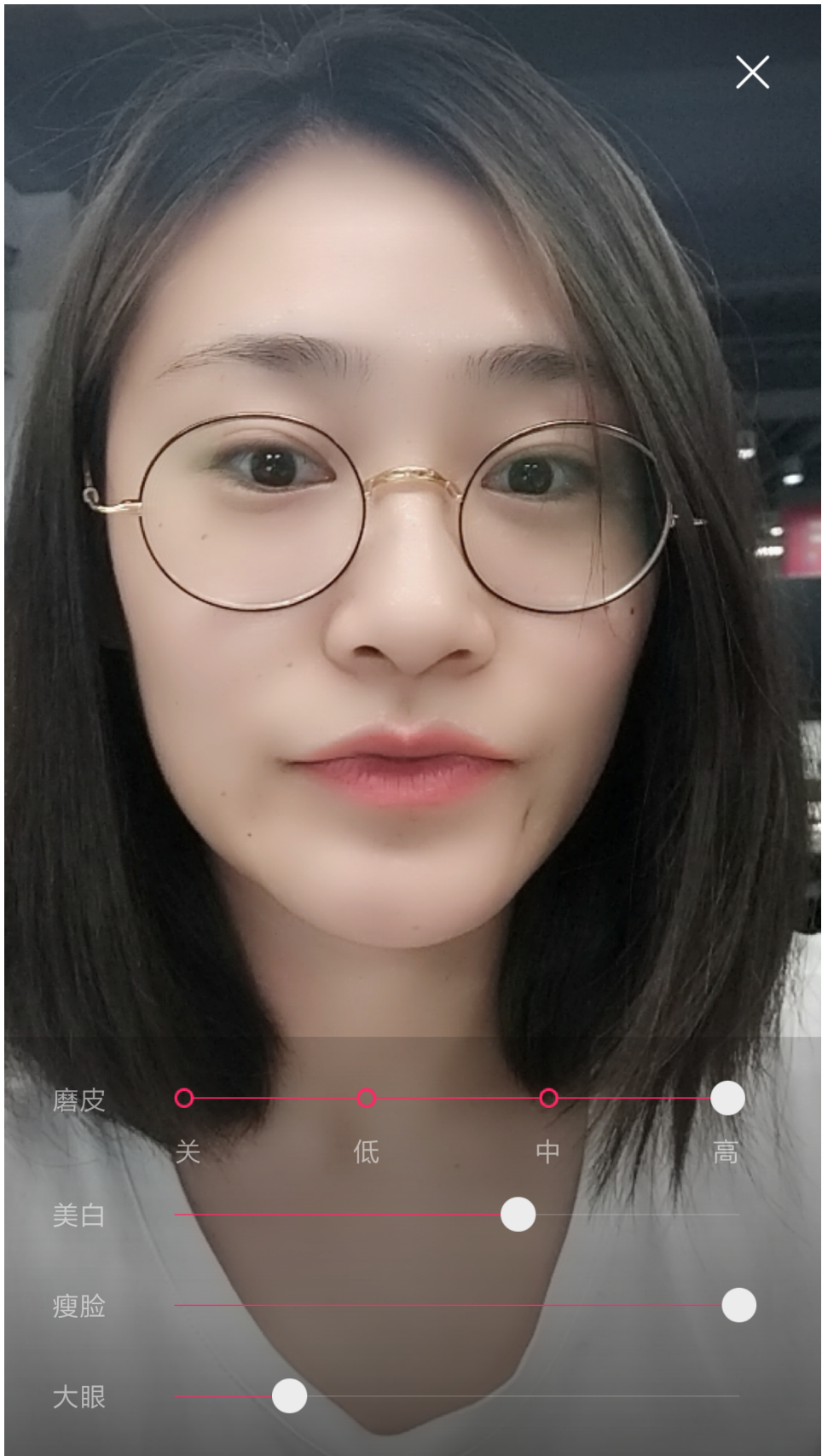
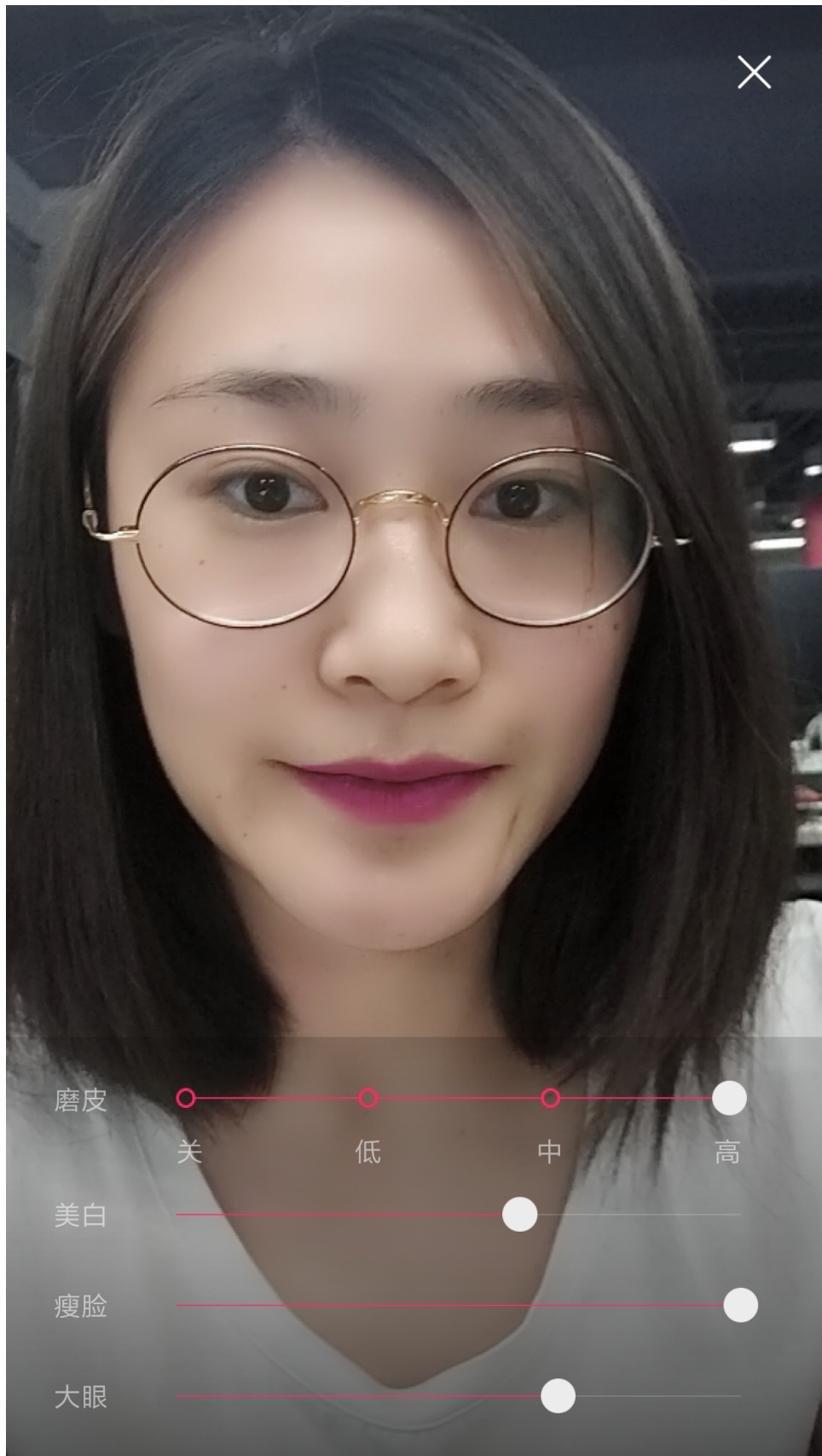
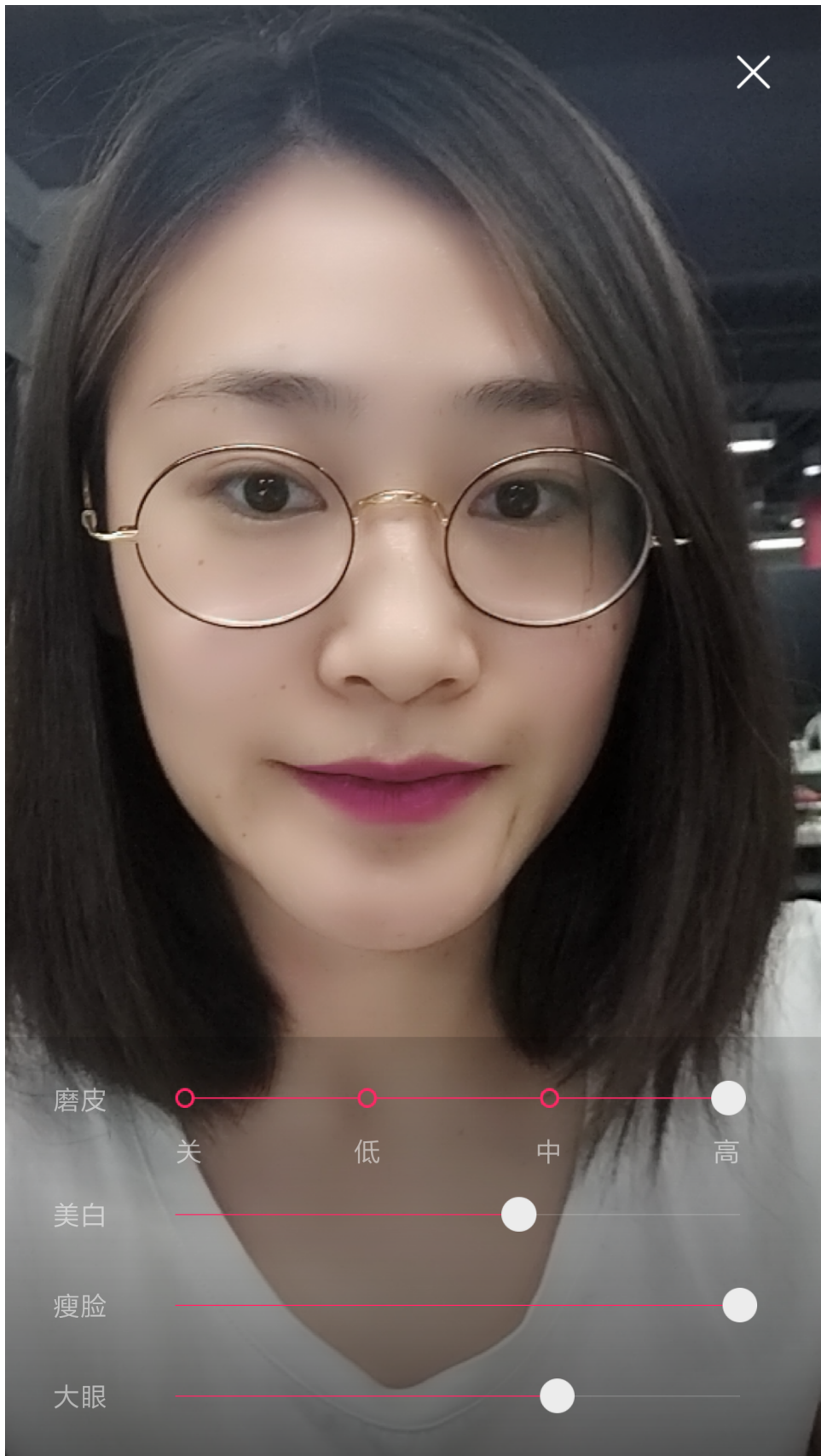
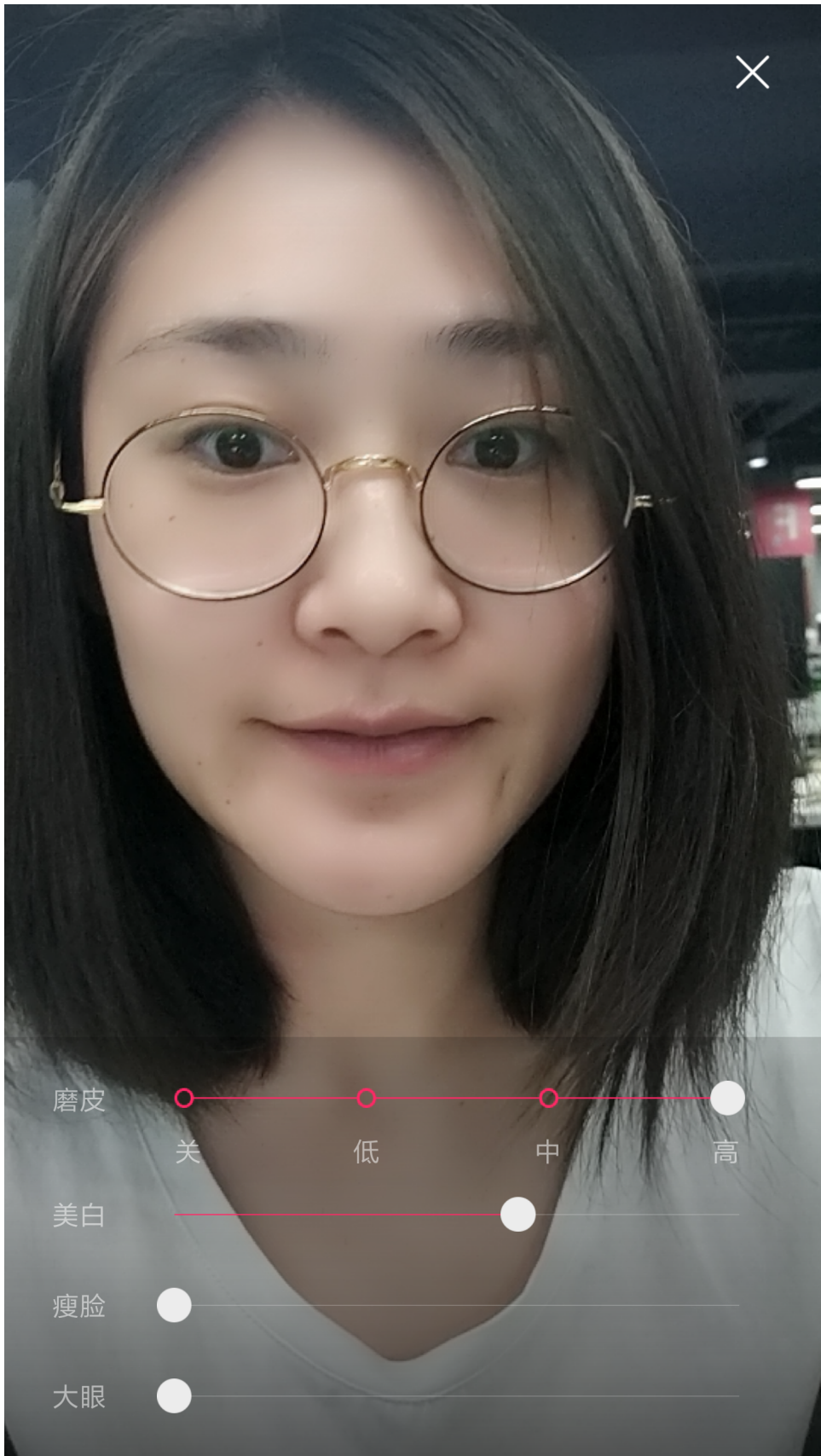
技术演进历史



功能实例



功能实例



A large light red circle and a smaller blue circle overlap on the left side of the slide. The text 'PART 2' is centered within the red circle.

PART 2

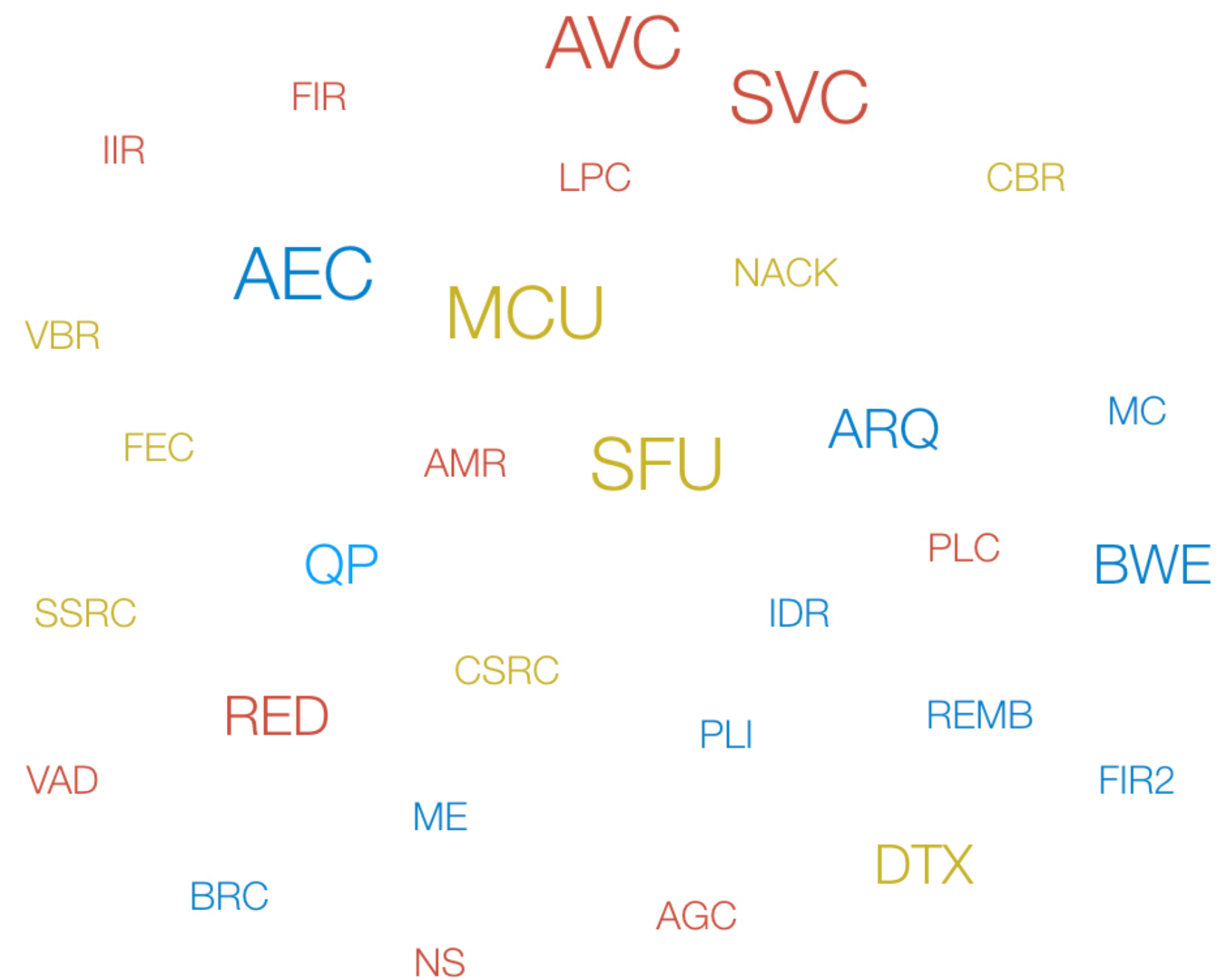
实时通话技术

音视频业务背景、[实时通话技术](#)、推流 / 播放技术、视频编辑技术

实时音视频 - 场景

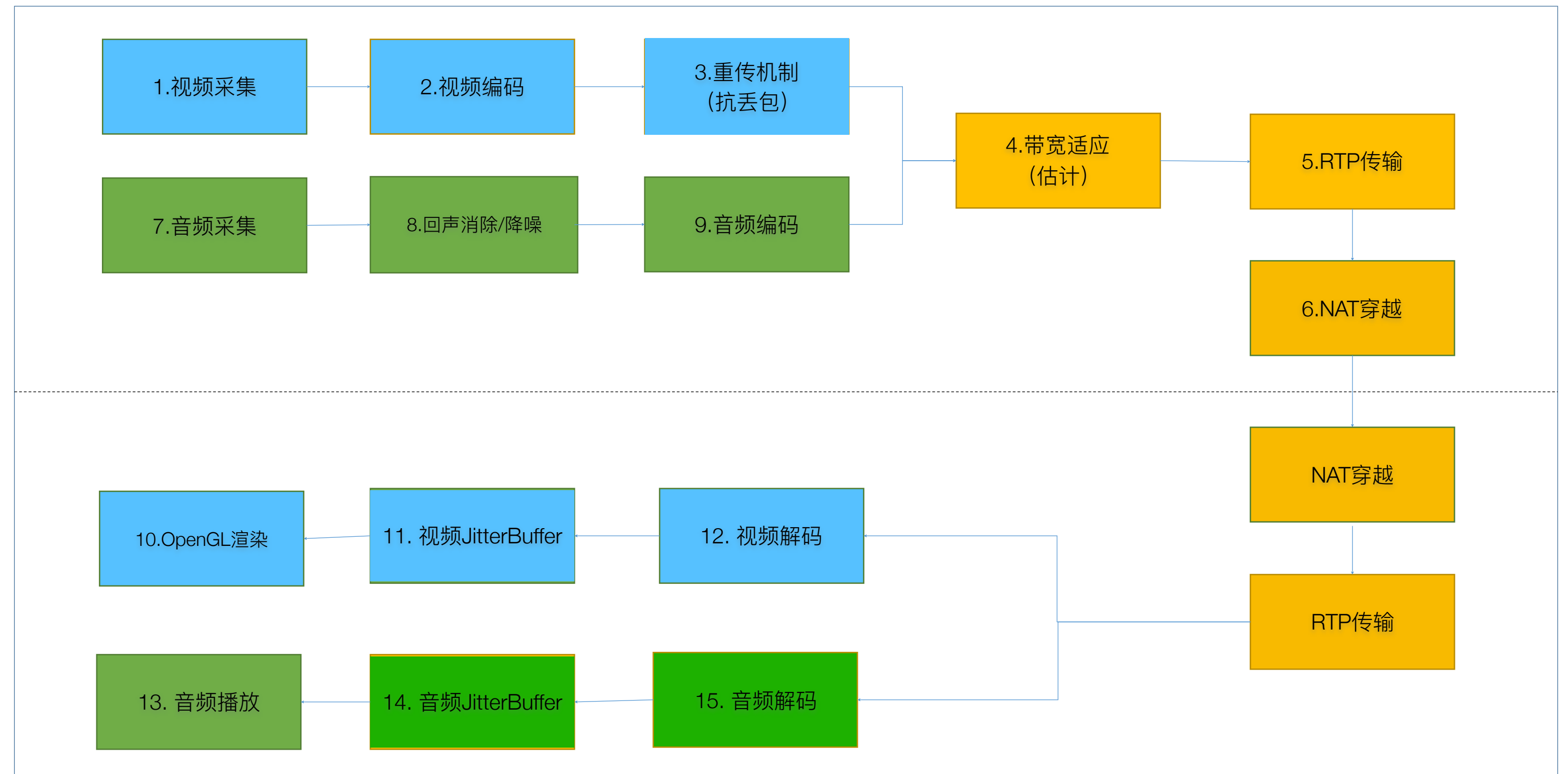
- 音视频最复杂系统
- 适合场景
 - 实时数据源
 - 双向音视频交流
 - $\text{delay} < 0.5\text{s}$

- 相关技术



技术方案 - 2人连麦系统

- 小米直播最早 (2016 / 05) 一批支持主播连麦
- 针对性优化远程协助功能, 降低操作延迟

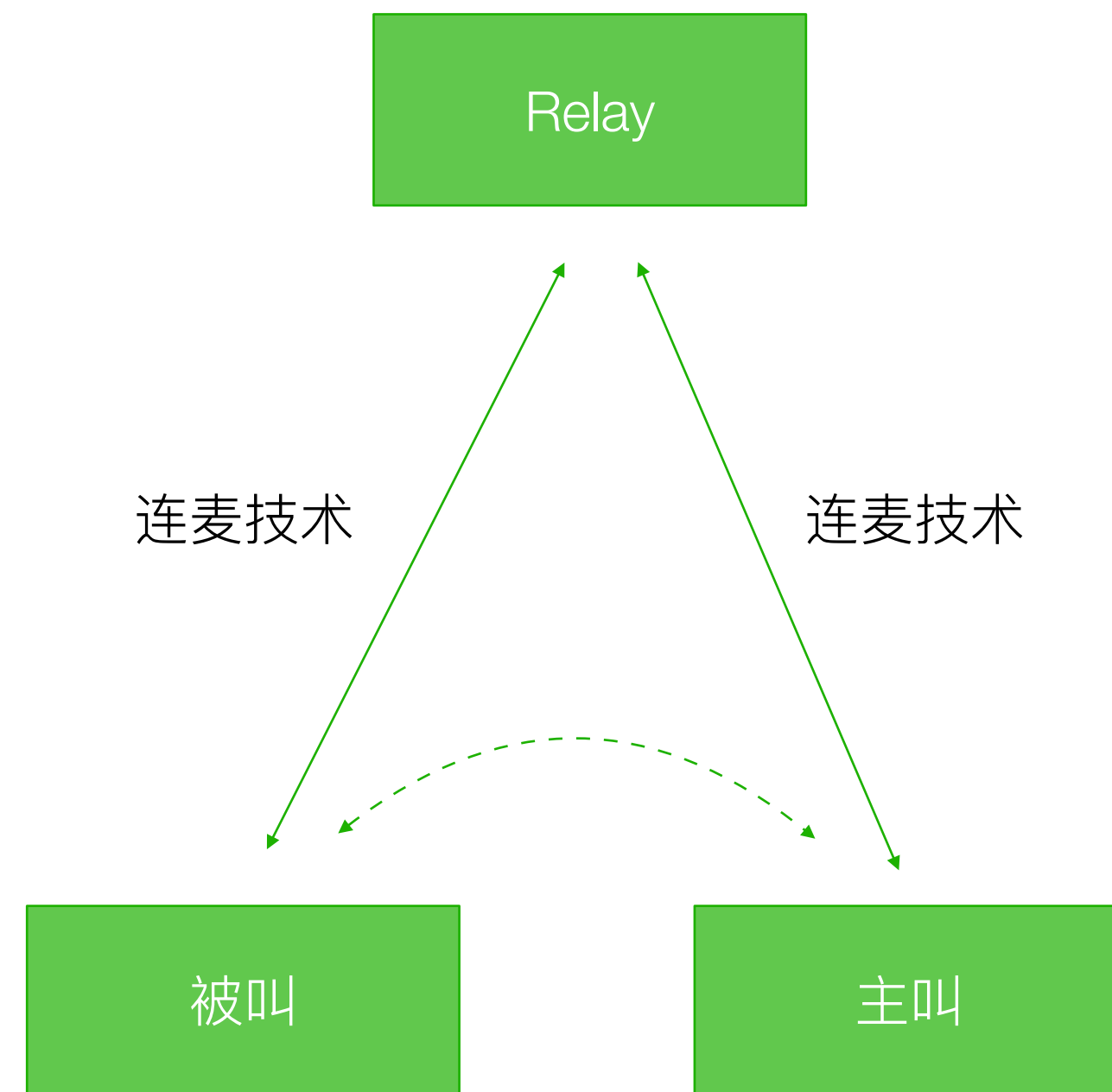


技术方案 - 多人连麦技术系统

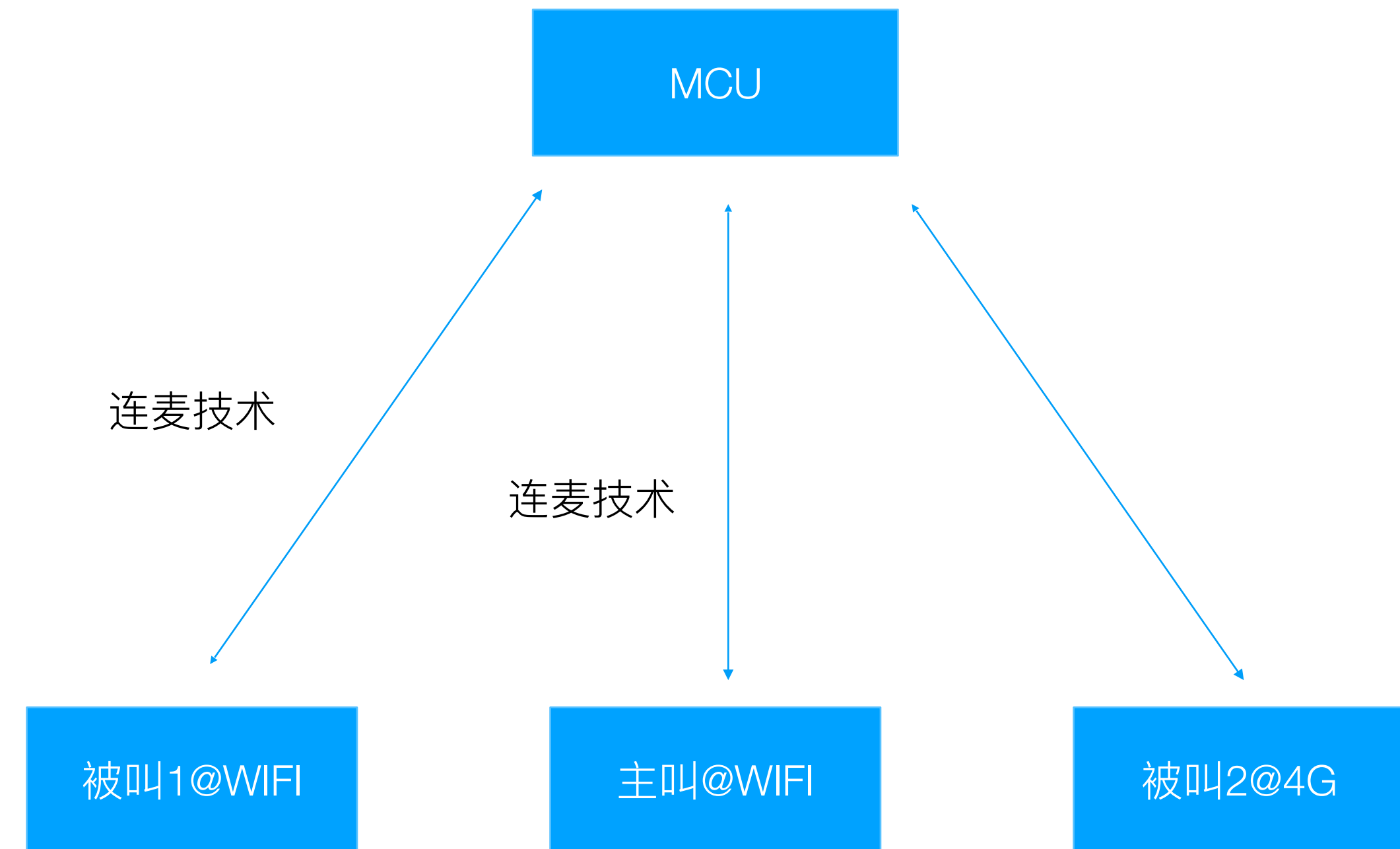
- 多人高扩展性 (**SVC**) MCU服务开发
 - 12人音频房间支持5000人
 - 6人视频房间支持3000人
- 从技术上支持狼人杀、电台等功能
- 而传统方案MCU只能支持~150路
 - 10人会议：编码10路、解码10路

技术方案 - 2人 vs 多人技术差别

2人连麦架构图



多人连麦架构图

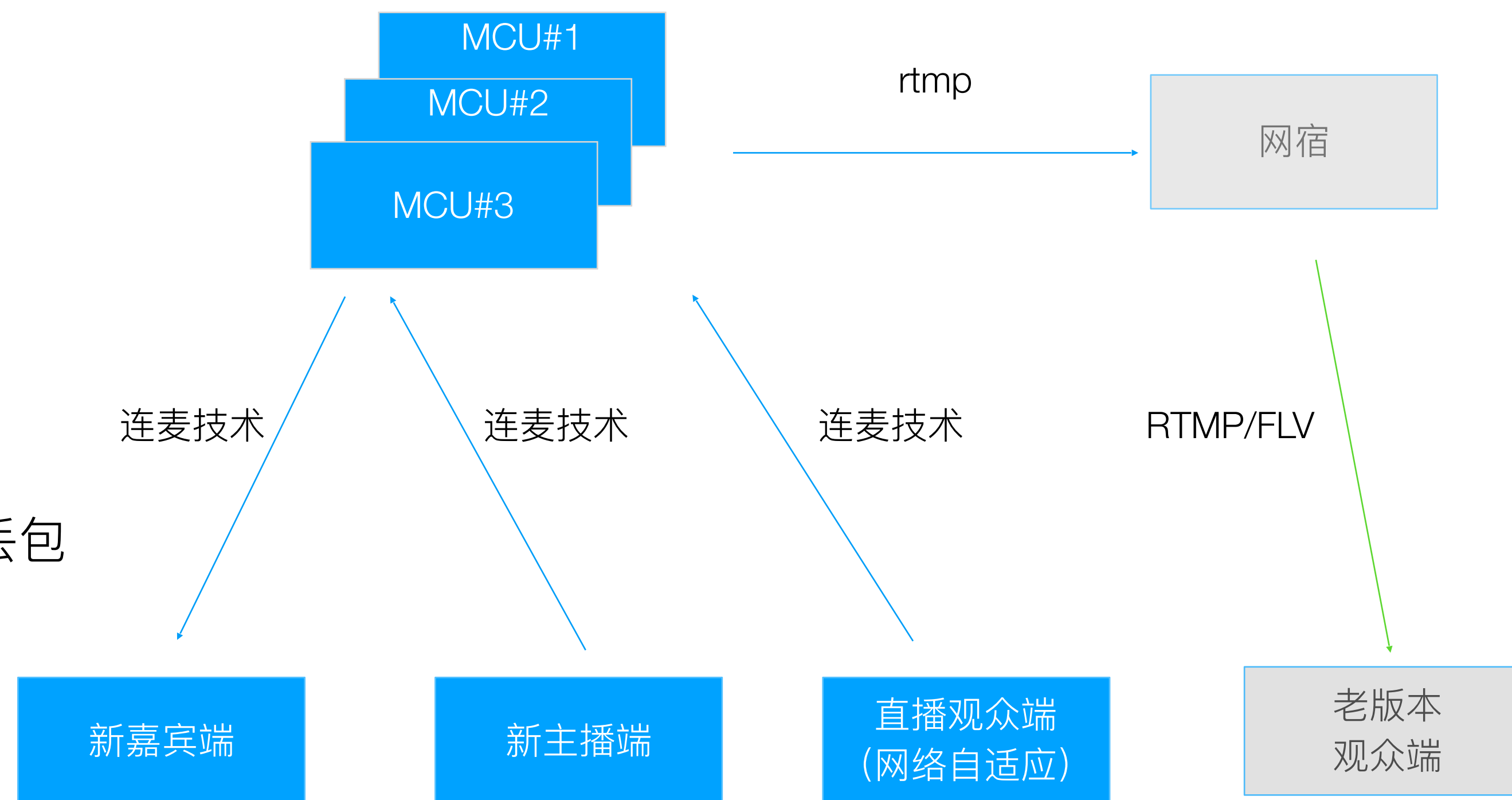


方案示例 - 电台方案（多人连麦直播）

● 融合实时通话与直播技术

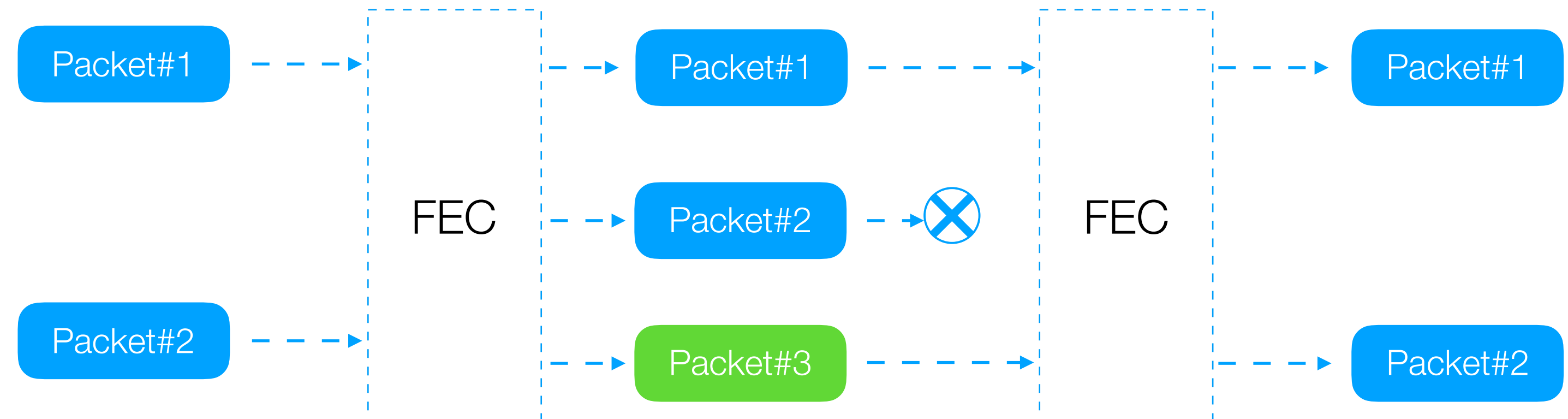
● 服务器推流

- 全系统48k立体声支持
- 降低客户端CPU消耗
- 主播端网络状况适应：更清晰画质、更强抗丢包
- 推流转连麦 无切换



技术难点 例1- 丢包问题

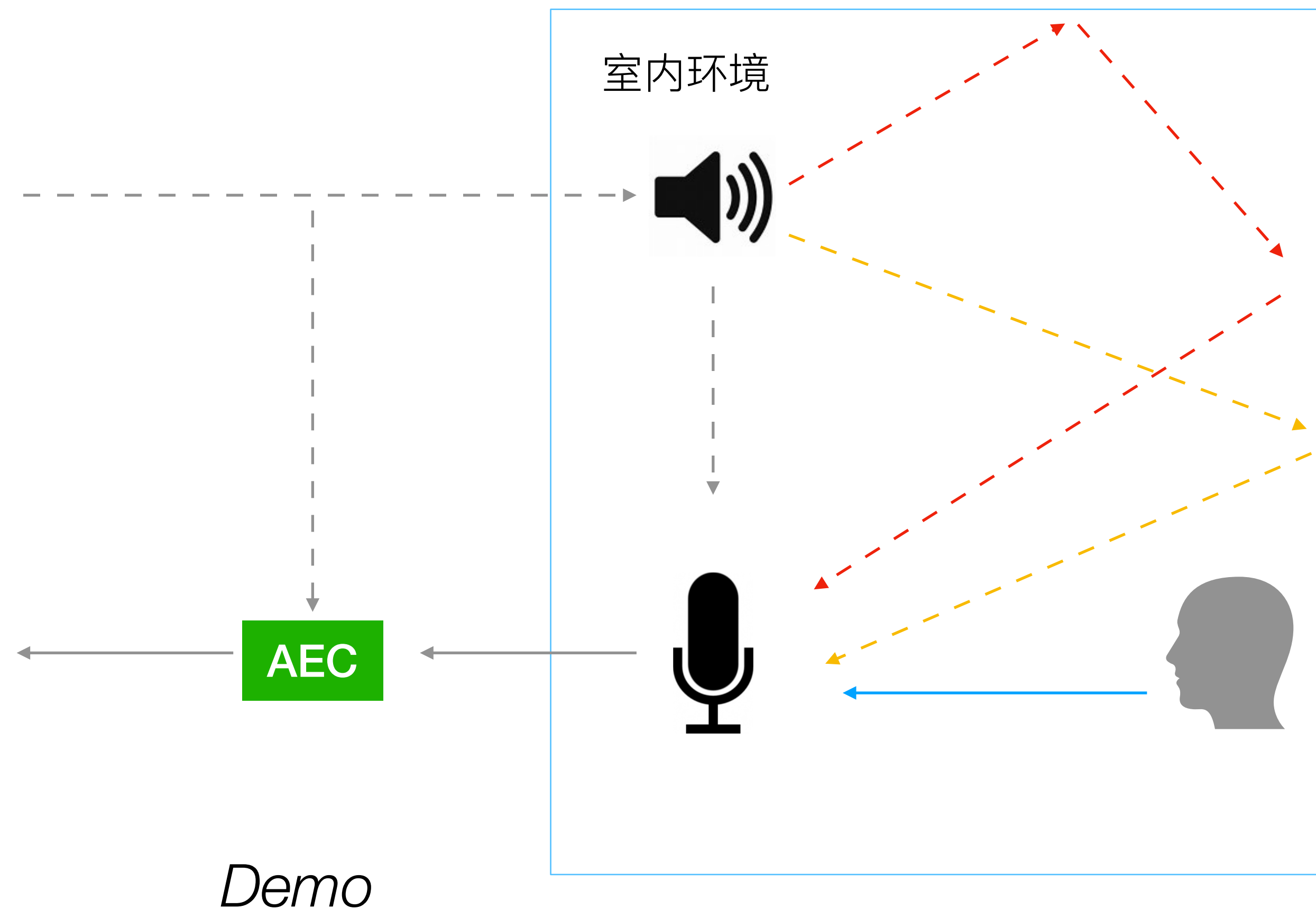
- FEC - 前向纠错 (Audio & Video)
- ARQ - 自动重传 (Video)
- FIR - I帧请求 (Video)
- RED - 动态冗余包 (Audio)
- PLC - 丢包补偿 (Audio)
- BWE - 带宽估计 (Video)



技术难点 例2 - AEC（回声问题）

- 基本问题

- 音频最困难问题之一
- 与通话环境相关性强
- 实时音频 质量决定因素

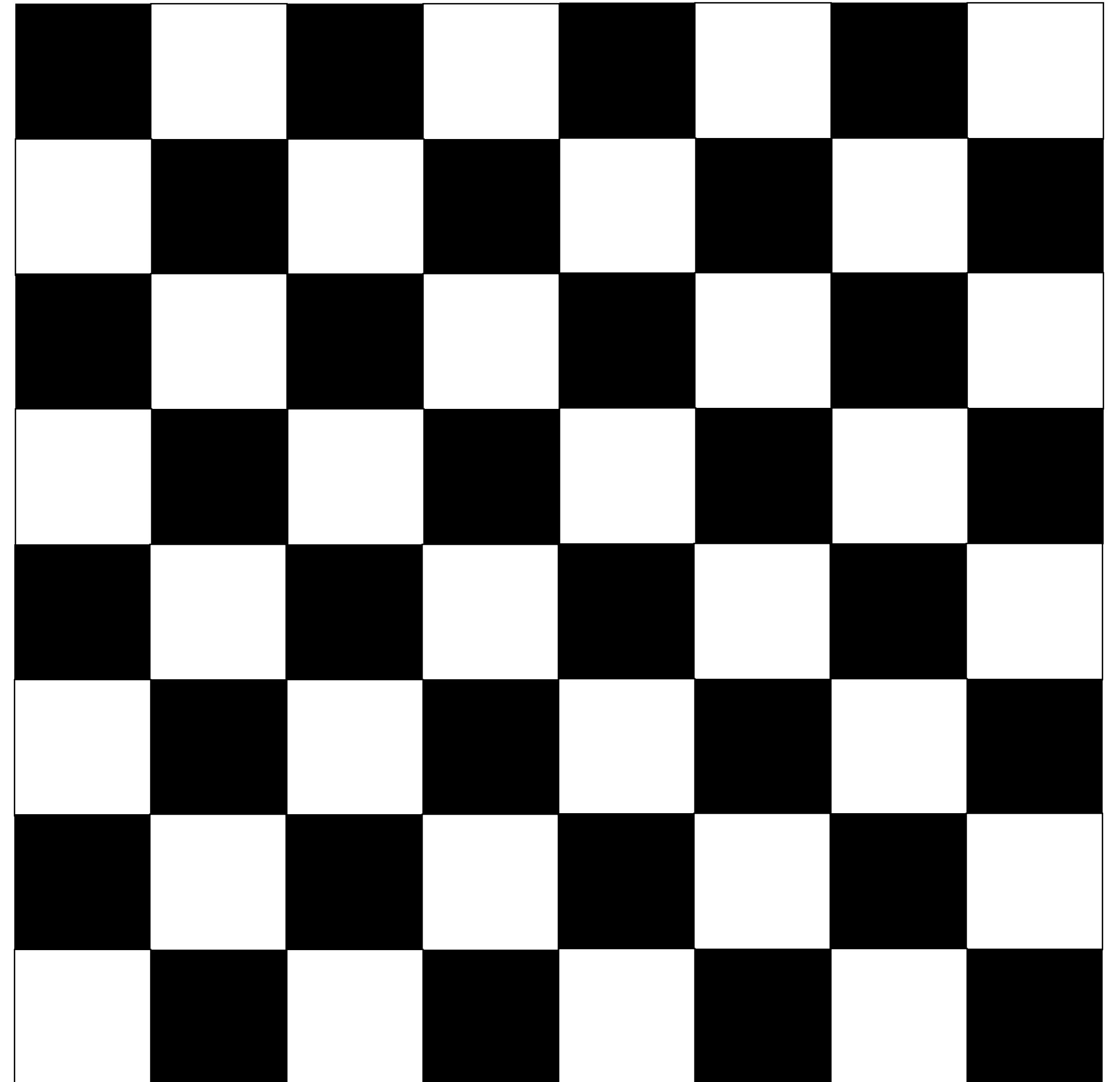


技术难点 2 - AEC（回声消除）

- 为什么不用硬件AEC?
 - 没有统一API
 - 效果不一致
 - 很多机型不提供
- 基于机器学习AEC算法
 - Android机型零适配
 - 效果赶超硬件（支持48k，硬件支持16k）
 - 单讲正确率 96%、双讲正确率85%（Mi6测试结果）

技术难点 例3 - Render (缩小)

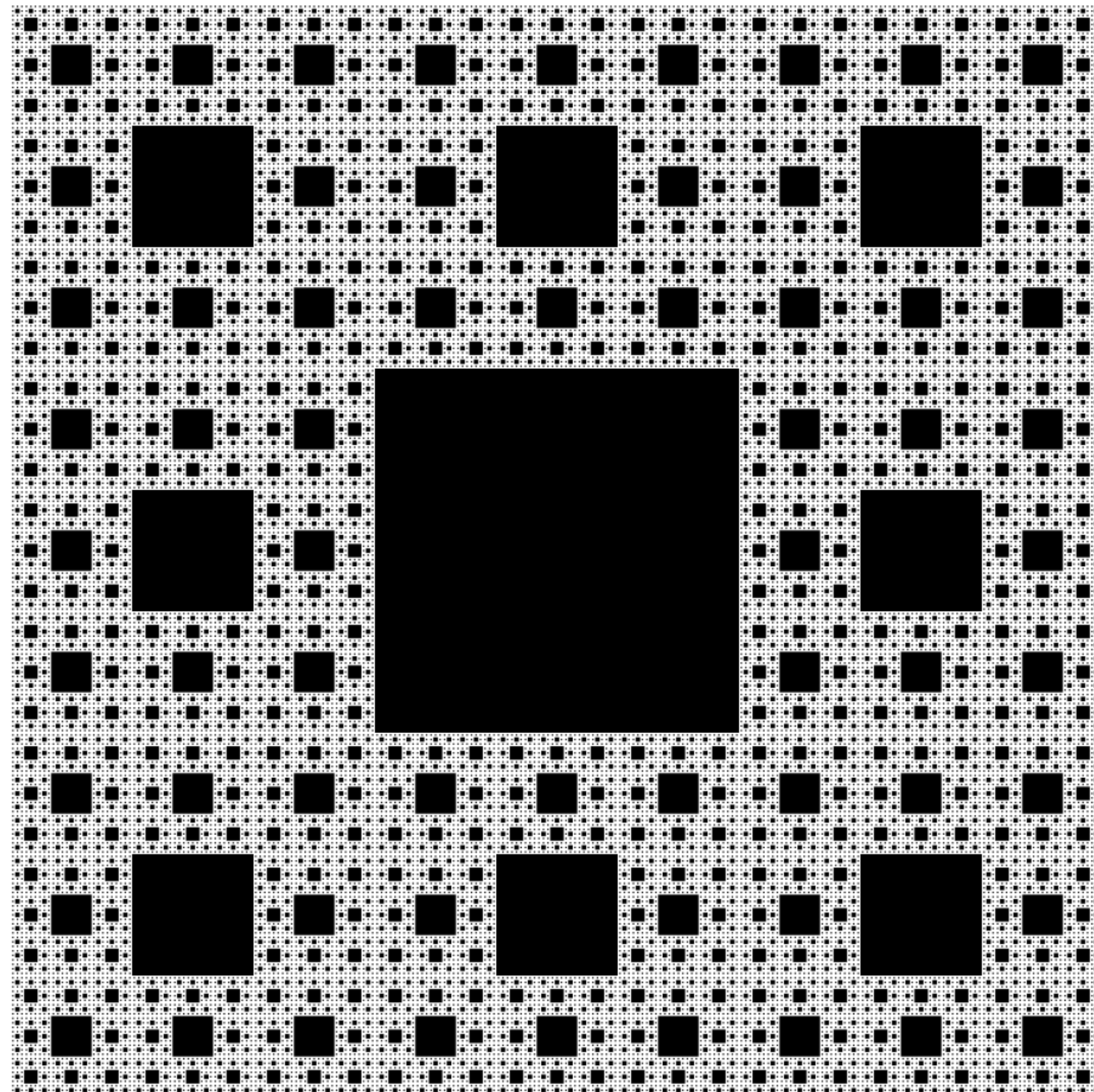
- 原因探究
 - 右图如何缩小 $1/2x$?
- 放大和缩小失真原因不完全一样



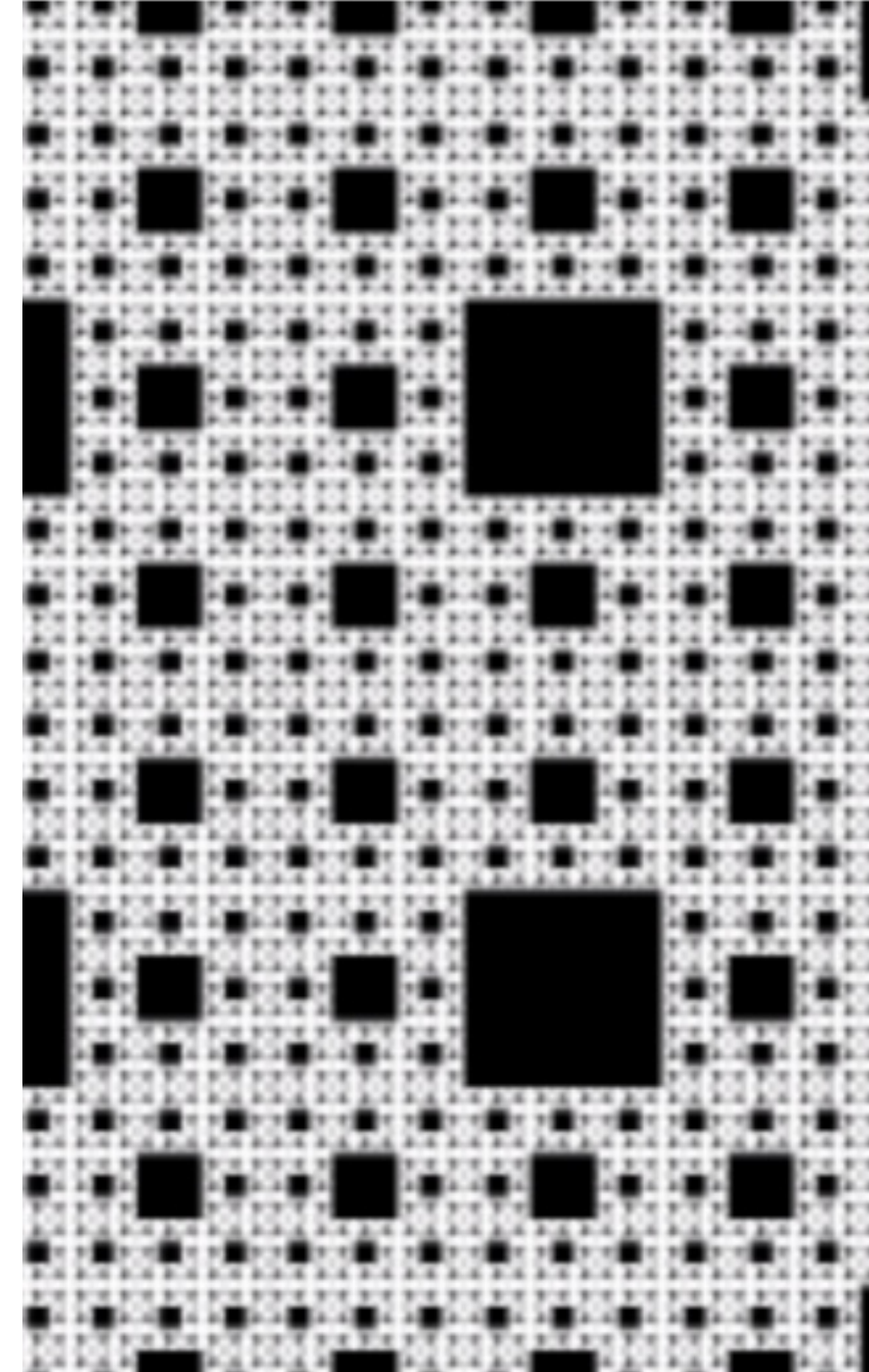
技术难点 例3 - Render 测试方法

- 锯齿很难察觉

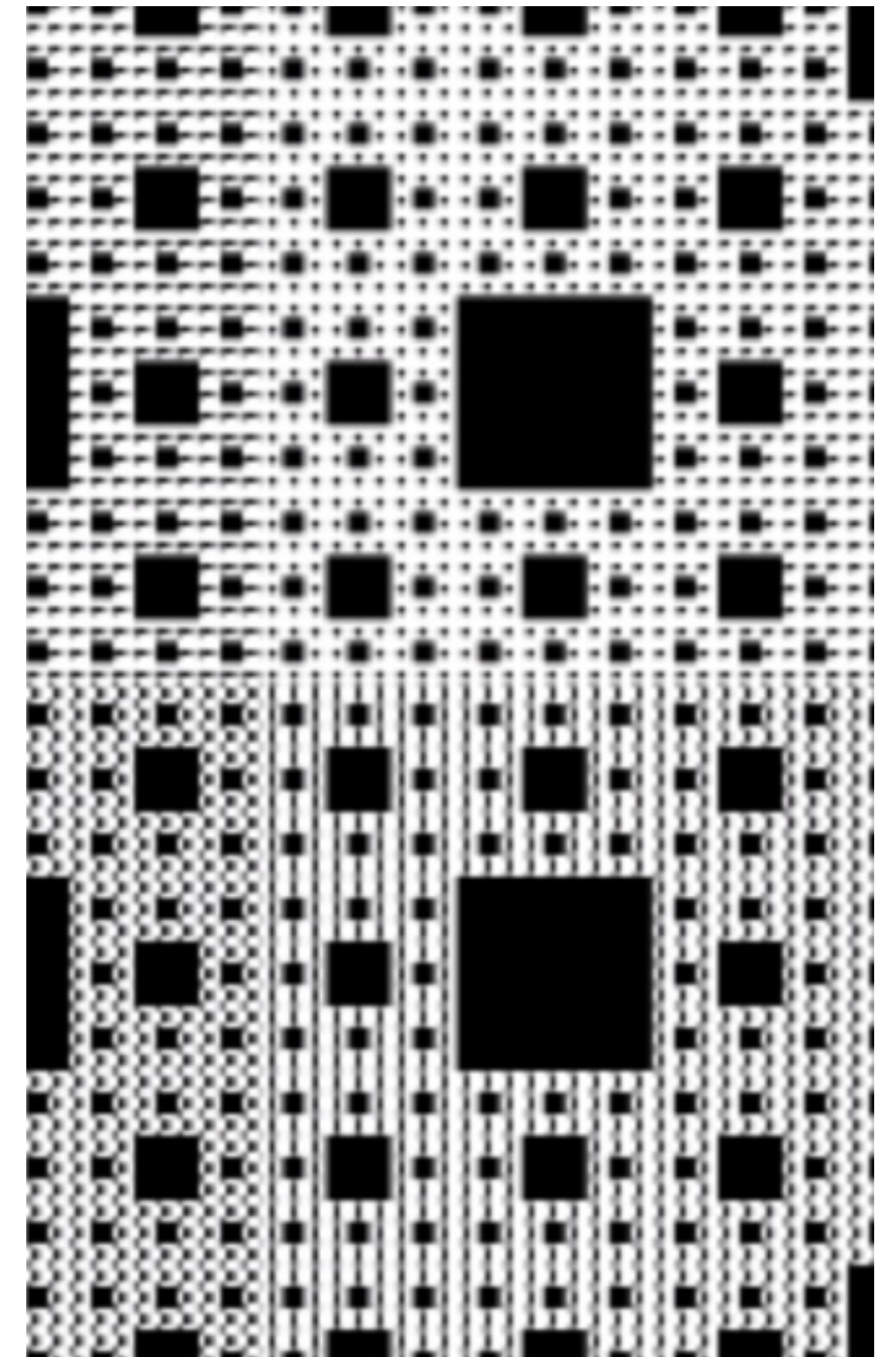
输入：



正确输出：



问题输出：



A decorative graphic on the left side of the slide features a large light green circle and a smaller blue circle that overlaps its bottom right edge.

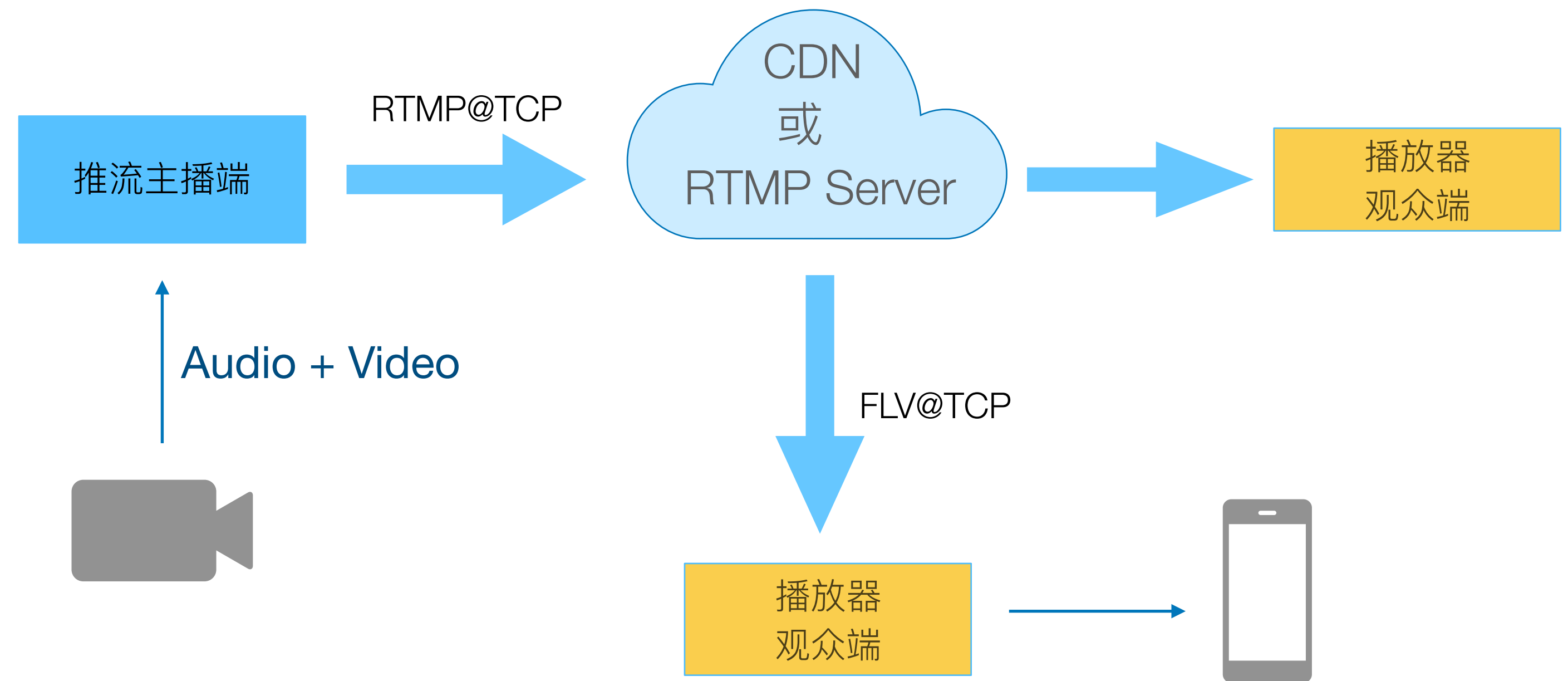
PART 3

准实时音视频（推流 / 播放技术）

音视频业务背景、实时通话技术、推流 / 播放技术、视频编辑技术

准实时音视频 - 客户端推流架构

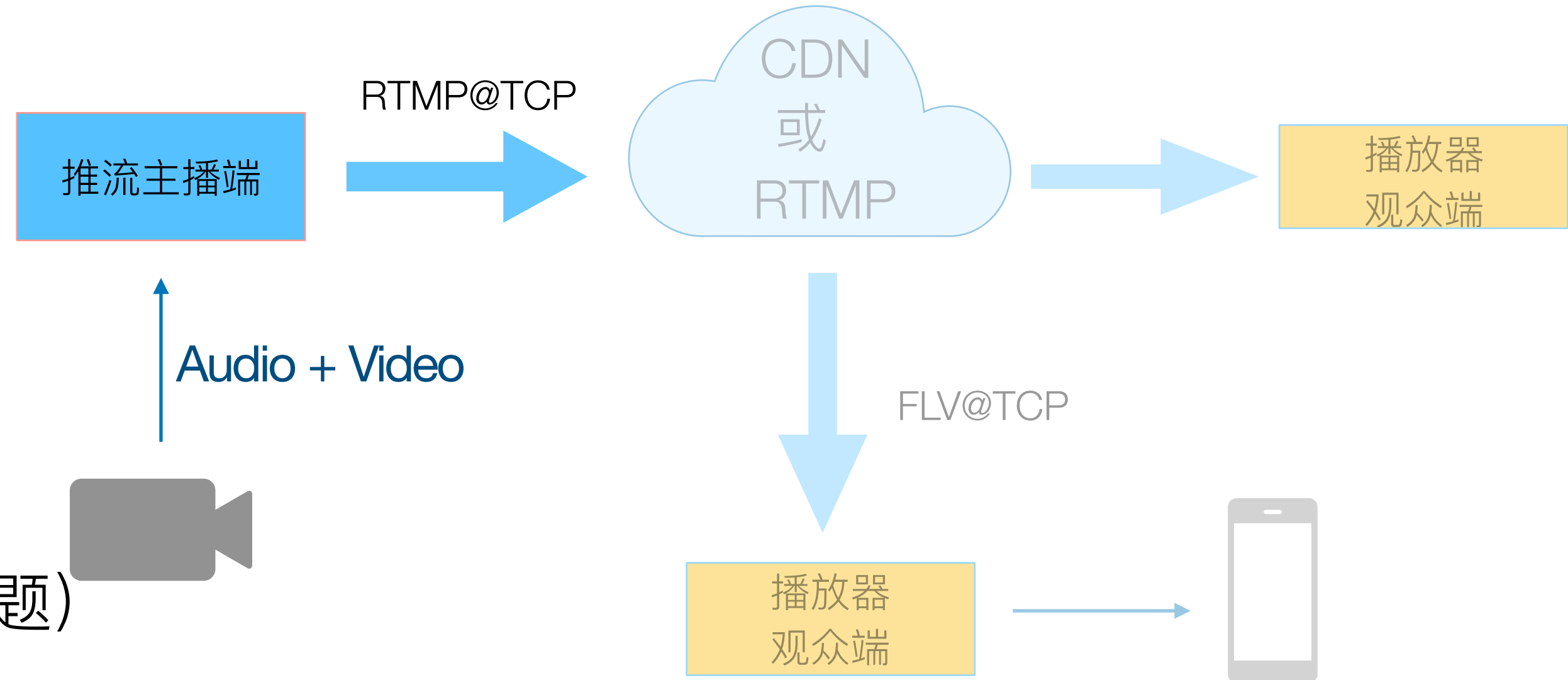
- 适合场景
 - 实时数据源
 - 单向音视频交流
 - delay < 5s



准实时音视频 - 客户端推流

- 推流相关技术

- FLV / RTMP
- TCP带宽估计
- H264 / AAC 编解码
- CDN
- 重连策略
- RTMP server
- I帧控制
- 镜面反转问题 (产品问题)

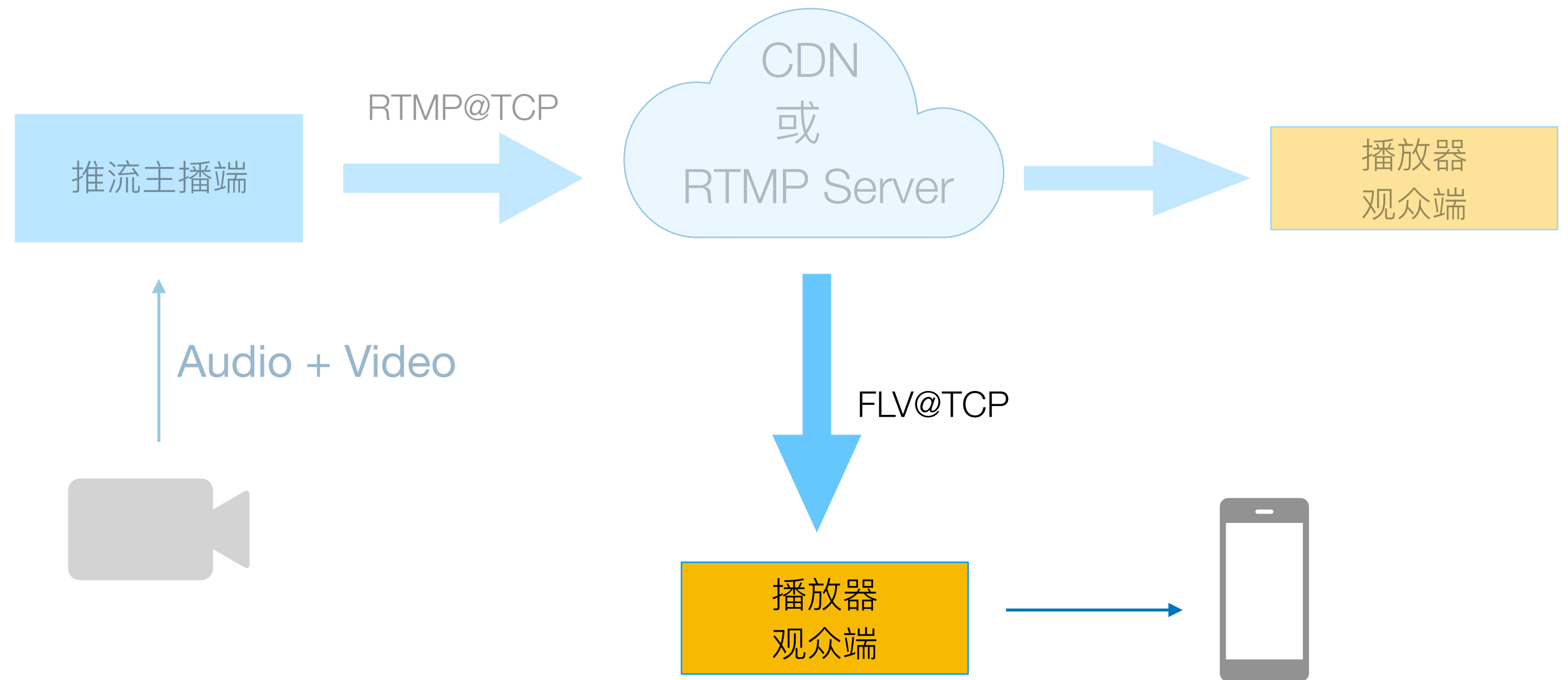


准实时音视频 - 客户端拉流、播放

- 播放相关技术

- 快播 / 慢播
- 快速重连策略
- FLV文件格式

- 唇音同步
- 缓存控制

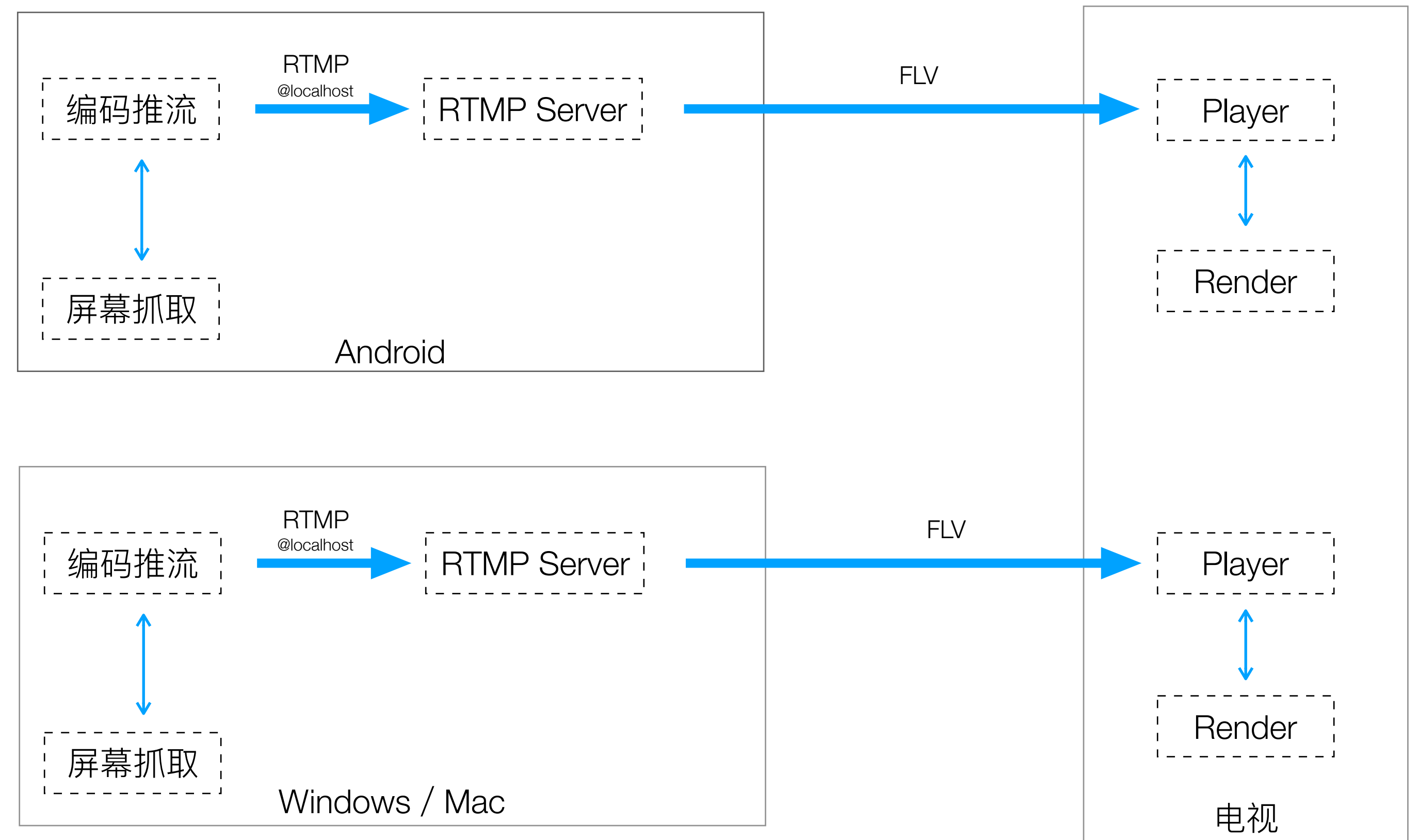


方案示例 - 某投屏项目

- 功能介绍

- 相关技术

- Delay ~ 0.4s
- 支持多个设备同时投屏
- 公司网环境（不要求同一网段）
- 做到：视频 / 音频 流畅播放



播放 - 重要功能点

- 秒开功能 / 首帧时间
- 播放器Seek响应时间
- 边下边播功能、预加载功能
- 硬件解码（4K视频支持）
- 文件格式支持（mp4/flv/m3u8/...）
- 媒体格式支持（H264/H265）
- 准实时流（直播流）播放功能

播放 - 重要功能点

- 播放器支持

- H265 / H264支持
- 音频110%速度无感快播
- 最高快进8倍速
- 播放器秒开 & 快速Seek
- 内存磁盘Cache
- 边下边播
- 全平台



播放器 - 文件格式介绍

- 格式对比

- FLV (FLASH VIDEO)
- M3U8 (MP3 URL @ UTF-8)
- MP4 (MPEG-4)

	FLV	MP4	M3U8 (TS流)
秒开速度	✓ ✓	✗ ✗	✓ ✗
Seek速度	✗ ✗ (需要扩展)	✓ ✓	✓ ✗
支持动态 分辨率切换	✓ ✓	✗ ✗	✓ ✓
边下边播	✓ ✓	✓ ✗	✓ ✗
缓存到本地	✓ ✓	✓ ✓	✓ ✗



PART 4

视频编辑技术

音视频业务背景、实时通话技术、推流 / 播放技术、[视频编辑技术](#)

编辑器

- 基本功能

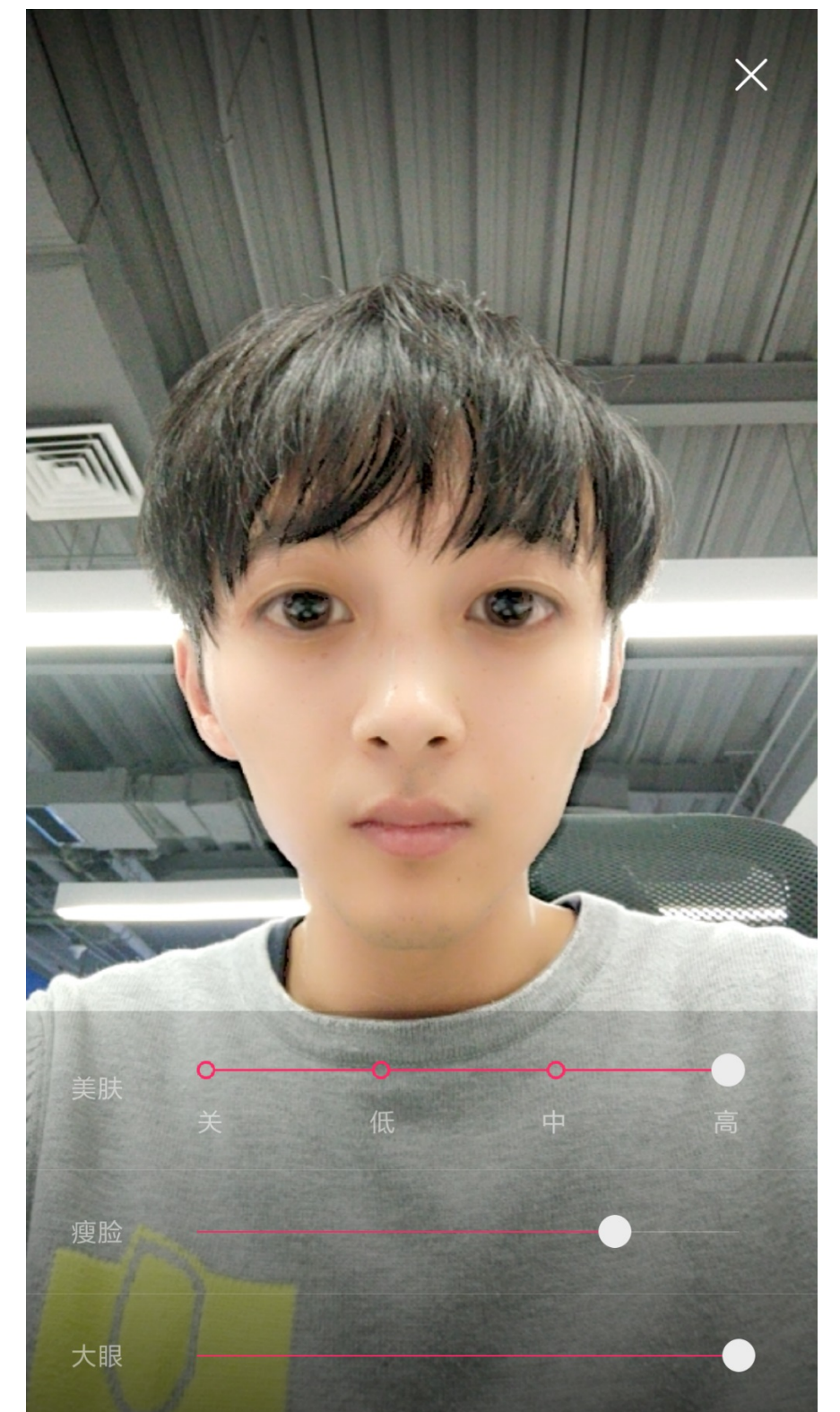
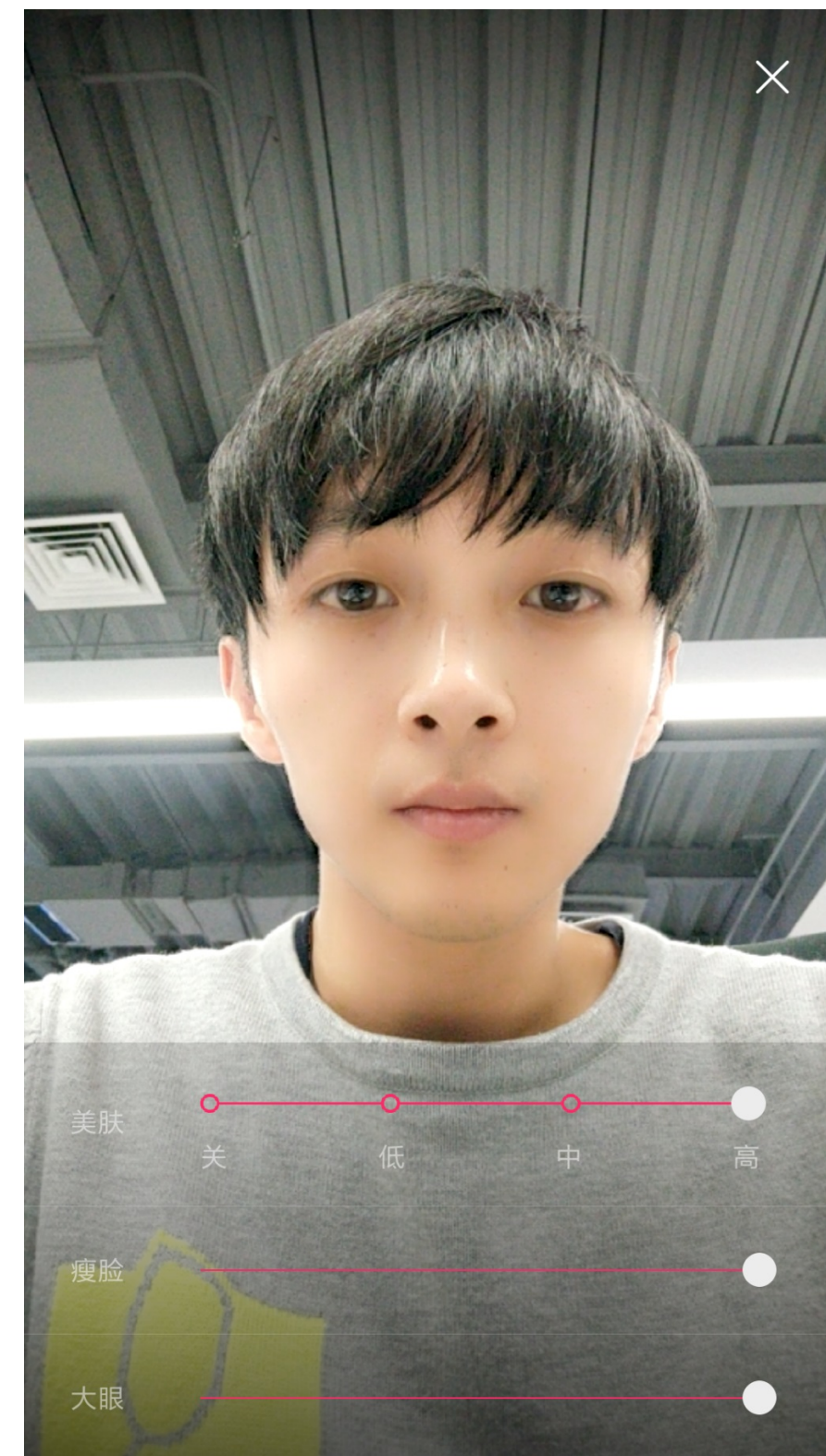
- 加混响
- 变声
- 加音乐、文字、图片
- 速度调节
- 截取时间段
- 合并文件
- 视频特效
- GPU加速

Demo(点击播放)

编辑器 – GPU图像处理

- GPU图像处理

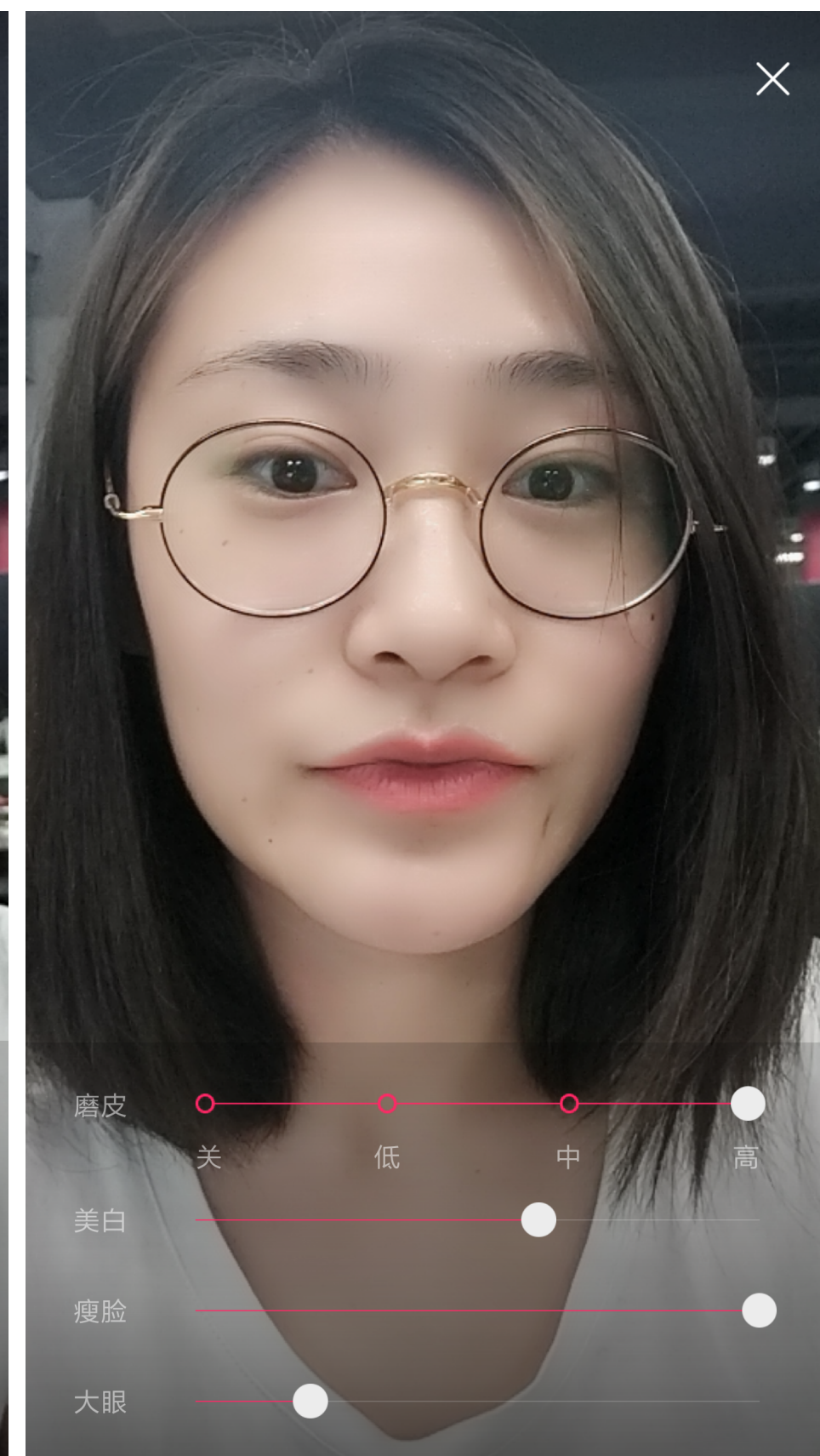
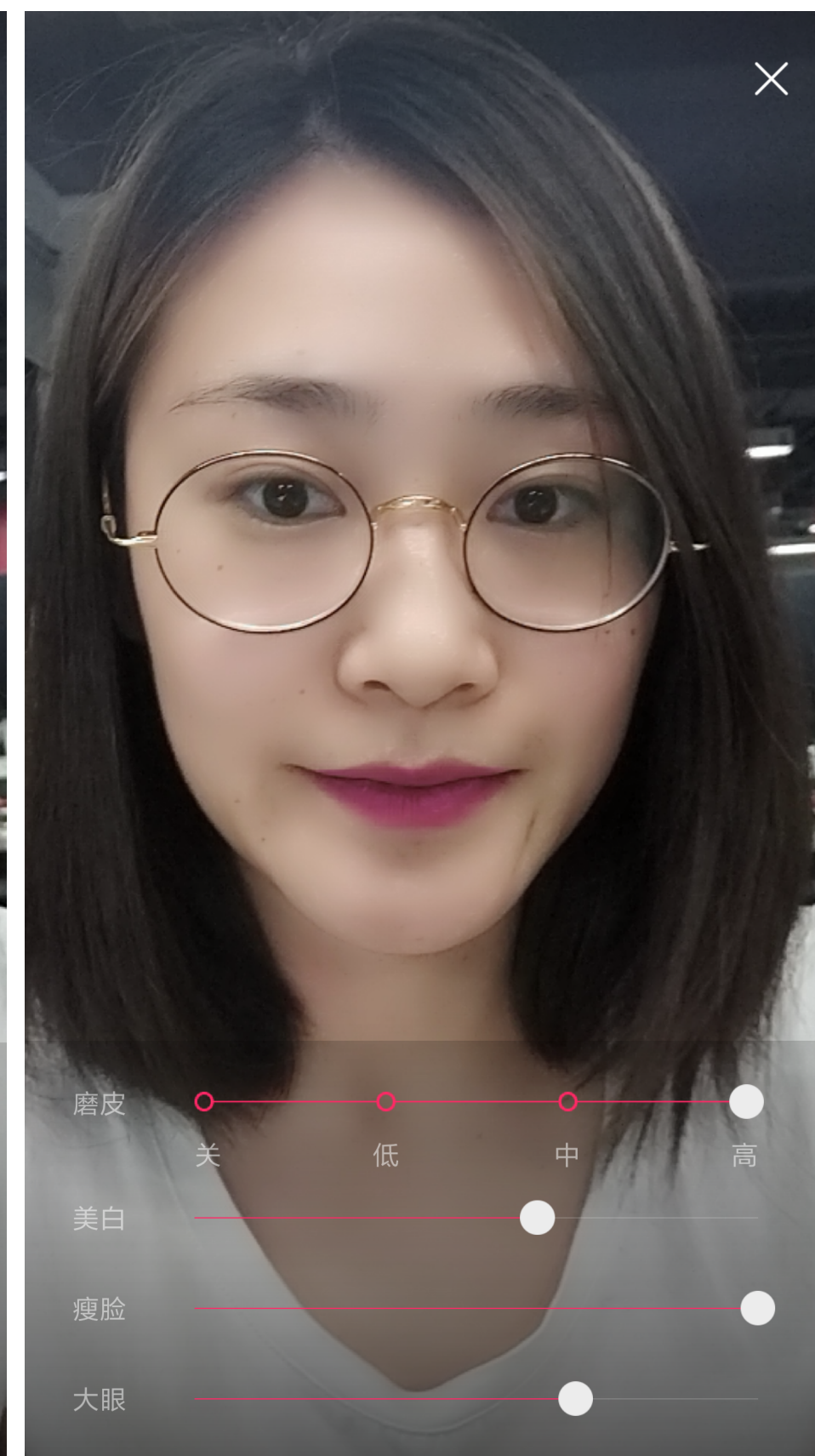
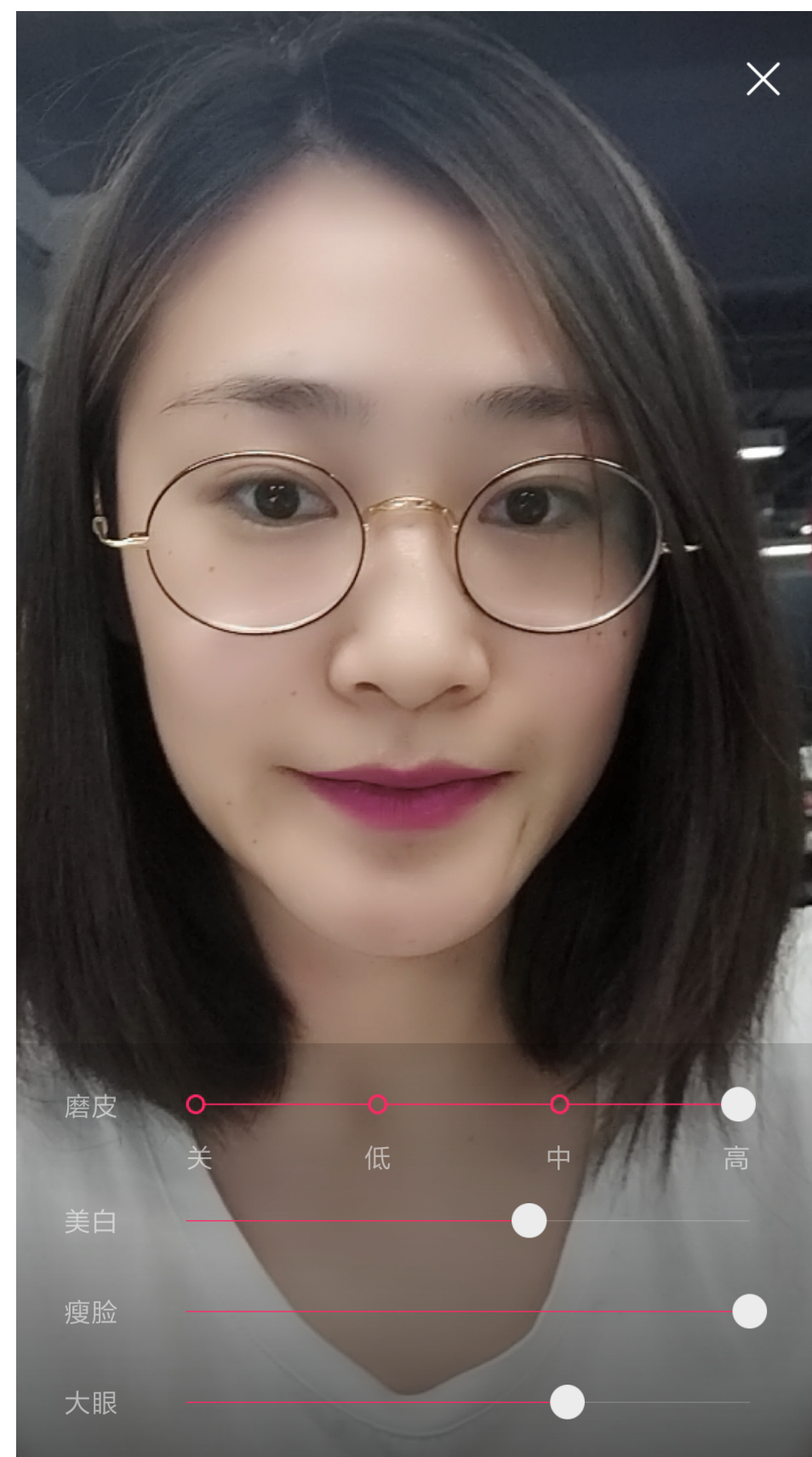
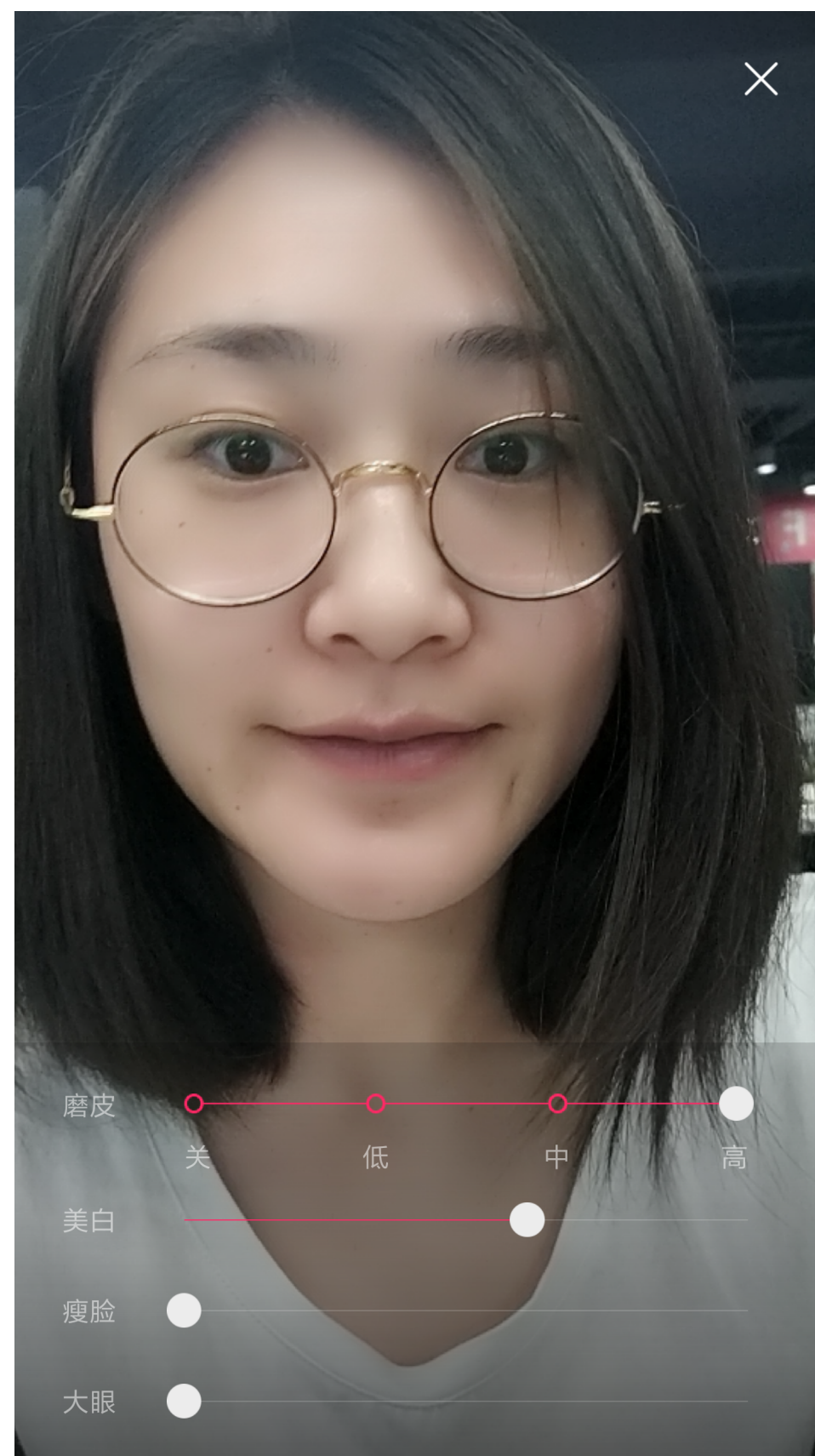
- 美颜（全局）
- 大眼瘦脸
- 动态贴纸
- 红唇



编辑器 – GPU图像处理

- GPU图像处理

- 美颜（全局）
- 大眼瘦脸
- 动态贴纸
- 红唇



编辑器 – GPU图像处理

- GPU图像处理

- GPU并行 & 矩阵运算
- GPUImage或者直接使用OpenGL
- GPU能力有限!!! 功耗问题!!!
- 硬件相关 - 需要全面机型测试
- Debug困难 (Image == Log)

```
// Vertex Shader
void main() {
    gl_Position =
        gl_ModelViewProjectionMatrix * gl_Vertex;
}
```

```
// Fragment Shader
void main() {
    gl_FragColor = vec4(0.2, 0.6, 0.8, 1);
}
```

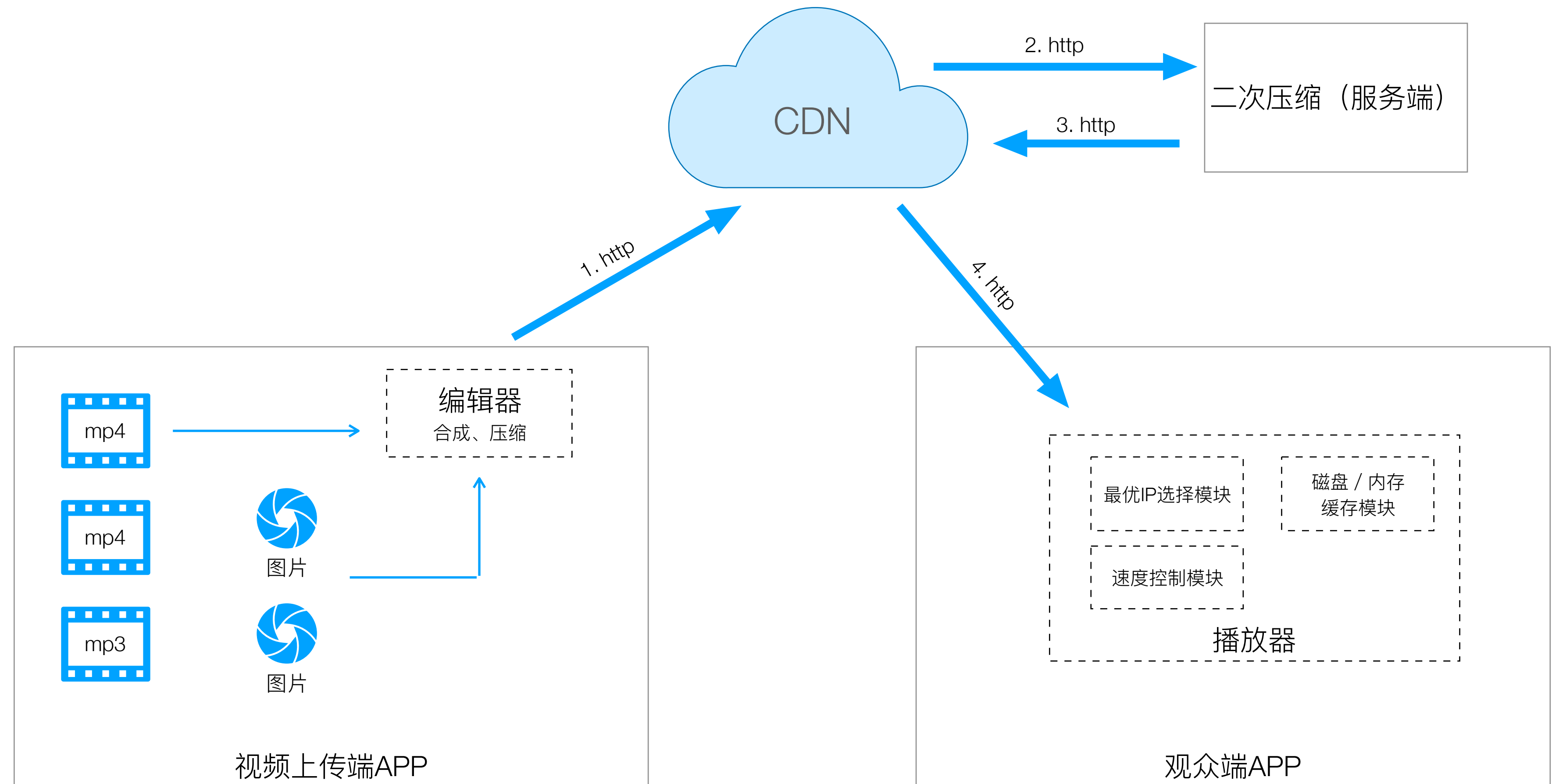
方案示例 - 视频编辑 上传方案

- 相关功能

- 压缩上传
- 编辑功能

- 实现技巧

- 同一文件格式快速合并
- 服务端二次压缩



全平台支持

- 全平台支持

- Android
- iOS
- Windows
- Linux
- Mac

- 例子

- Android / iOS: 各大 500+DAU app等
- Windows: 某投屏项目
- Linux: 导播台
- Mac: 某投屏项目

谢谢 Q&A



吴子敬 wuzijing@xiaomi.com