

交互式直播推流编码器的设计

赵文杰

学而思网校

演讲内容

- ▶ 直播编码器的使用场景
- ▶ 直播编码器在直播中的位置
- ▶ 直播编码器的典型架构
- ▶ 如何进行直播交互
- ▶ 交互式直播在教育上的独特应用
- ▶ 直播编码器在教育上的技术挑战

直播编码器的使用场景

娱乐
类

秀
场

游
戏

体
育

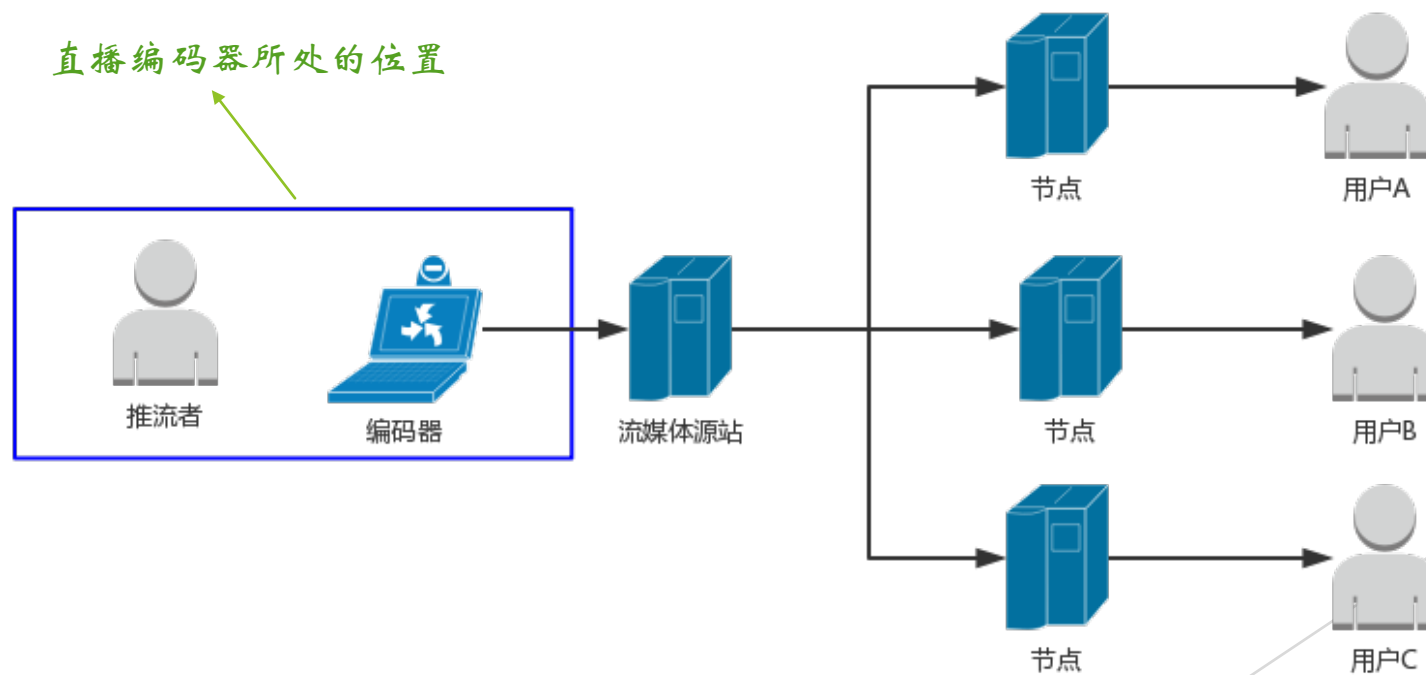
教育
类

在 线
教 育

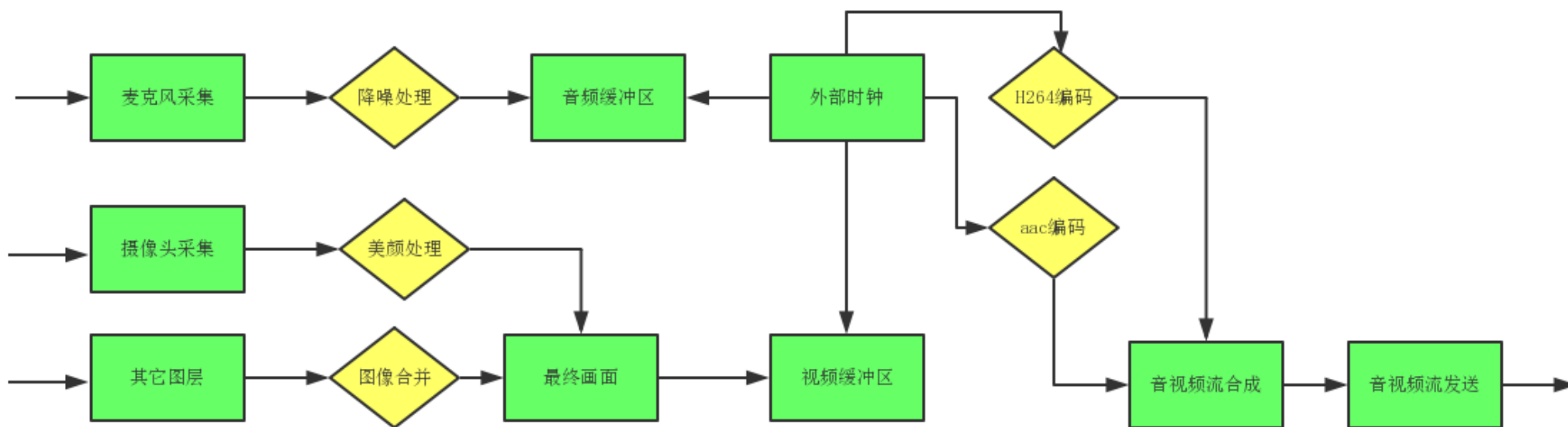
电商类

直 播
营 销

直播编码器在直播中的位置



直播编码器的典型架构

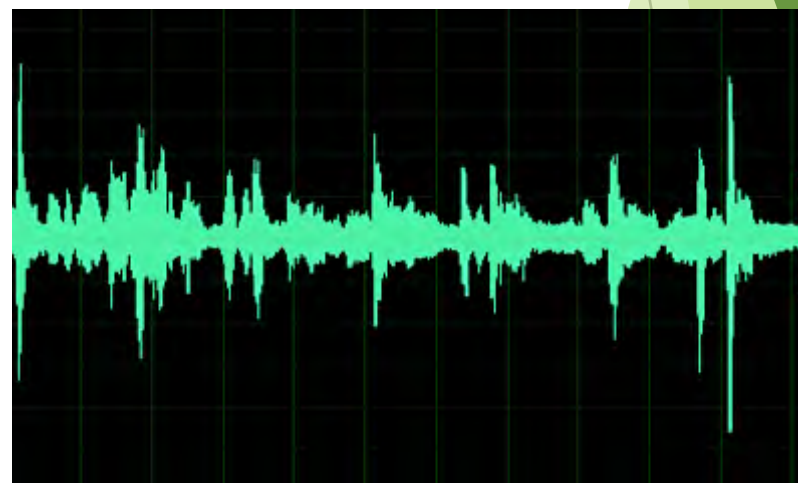
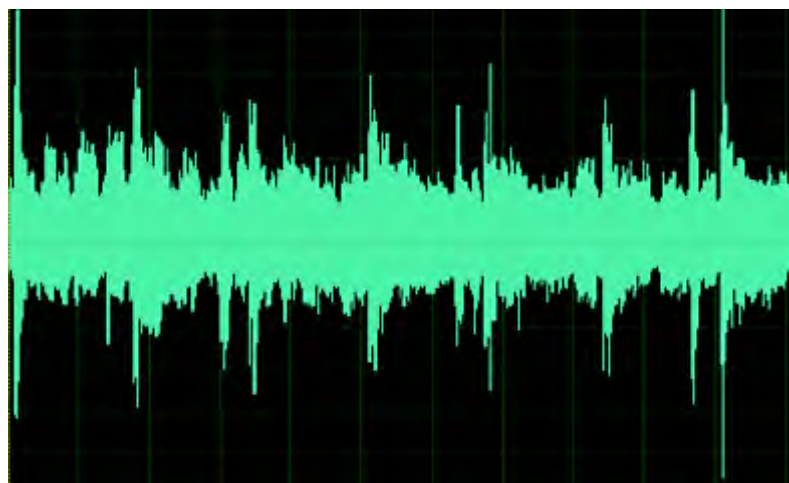


技术点分析 音频采集后处理

► PCM (Pulse Code Modulation) 数据的处理

目的是除去声音中电流声、环境噪声等，为终端用户提供更好的视听体验。

开源经典处理库可以参考webrtc的ns音频处理模块。



技术点分析

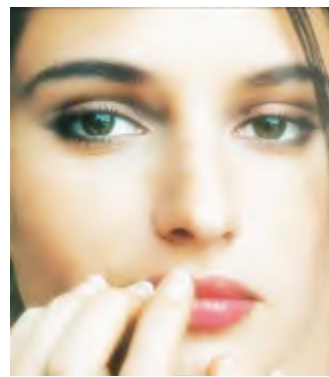
视频采集后处理

► RGB/YUV 的美化处理

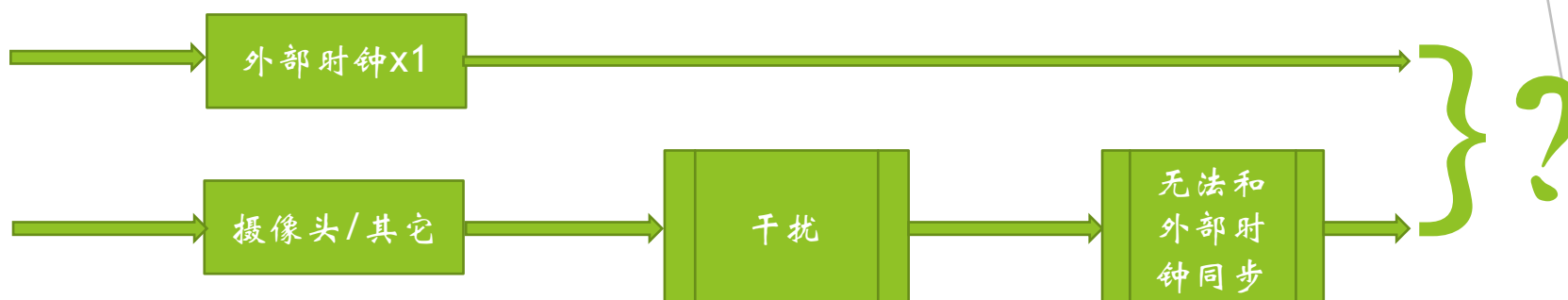
给摄像头或者其它直播画面增加一些特效处理，可以大大增强直播的趣味性。

由于图像处理对CPU的消耗很大，所以这个过程建议放到GPU里去做（如GLSL去处理图像），这个典型的一个开源库是GPU image。

或者通过第三方的视频美化软件来做（一般是虚拟摄像头）。



技术点分析 音视频缓冲区的目的



- ▶ 根据上图，由于采集会收到来自系统、硬件等方面的限制，可能会超出或者低于外部时钟指定的采样率。

解决方法就是增加缓冲区，来减少这种干扰，如果采样率较低，可以使用最后一帧来填充，如果采样率较高，可以丢弃一些旧帧。

技术点分析 音视频同步

以RTMP/FLV直播流为分析媒体，假设：

视频帧率：25 fps 音频采样率：44100，每帧音频包含1024个采样

则video_duration=1000/25=40ms audio_duration=1024/44100*1000=23.219954ms



技术点分析 音视频合成

以RTMP/FLV直播流为分析媒体

- ▶ DTS
Decoding Time Stamp
- ▶ PTS
Present Time Stamp
- ▶ CTS
Composition Time Stamp

那么rtmp/flv里的时间戳是什么呢?

DTS

如果H264包含了B帧呢?

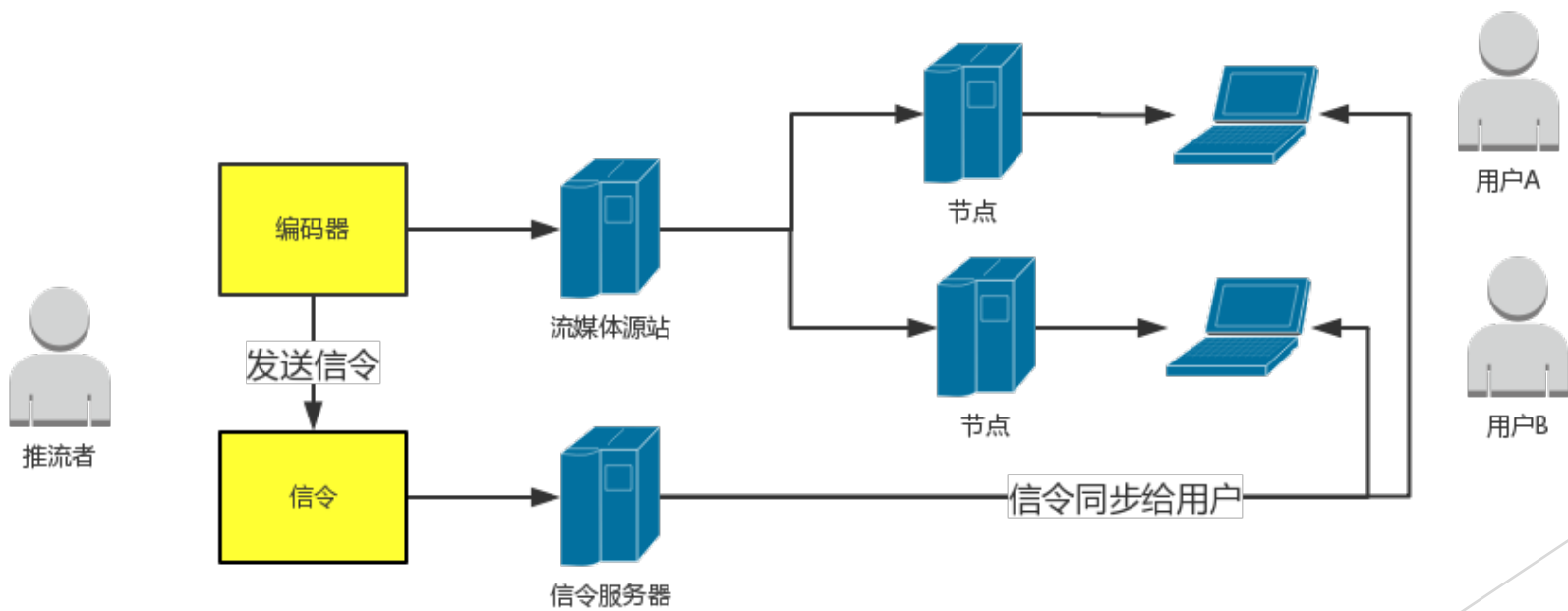
CTS

=

PTS



如何进行直播交互



交互式直播在教育上的独特应用

- ▶ 教师所见，学生即所得
- ▶ 录屏？
- ▶ 各种数学作图图形，画笔工具
- ▶ 辅助教学内容的展示
- ▶ 码率大小？



直播编码器在教育上的技术挑战和展望

- ▶ 低码率
- ▶ 内容丰富性
- ▶ 低延时
- ▶ 互动性
- ▶ 上传链路问题
- ▶ 3D素材的展示?
- ▶ VR教学?

谢谢大家！

赵文杰

学而思网校