M5 · 第 15 讲

配音、字幕与剪辑交付

把零散镜头变成一条能发的片子

阅读约 35 分钟

练习约 30 分钟

产出：15 秒成片

本节目录

开场：镜头都对，就是“不像一条片子”

本课目标

一、配音：先自然，再像谁

二、字幕：跟着呼吸走

三、音乐与音量

四、剪辑顺序

五、转场怎么用

六、导出参数

七、30 分钟练习

八、检查清单

九、延伸阅读

开场：镜头都对，就是“不像一条片子”

素材齐了，拼在一起却像 PPT：镜头之间硬切、配音忽快忽慢、字幕跳出来又消失。观众说不出哪里不对，但三秒就划走。

这一讲解决交付前的最后一公里：**声音、字幕、节奏、导出**。

本课目标

- 用三步把配音做得自然；
- 写出可读的字幕节奏；
- 用音量与转场把镜头粘起来；
- 按平台要求导出。

一、配音：先自然，再像谁

三条原则：

1. **短句**：一句话不超过 20 字，TTS 更自然；
2. **口语**：把书面语改成能说出口的话；
3. **标点控节奏**：逗号停半拍，句号停一拍。

流程：

- 先写口播稿，自己读一遍，改掉拗口处；
- 生成配音，调语速（每秒 4–5 字较自然）；
- 导出后听三遍：有没有吞字、断句错、情绪平。

不要一上来追求“克隆某个声音”。先让内容站住。

二、字幕：跟着呼吸走

规则	建议
每行字数	12–16 个汉字

规则	建议
每屏时长	1.5–3 秒
断句	按语义断，不按标点硬切
位置	底部安全区内，避开平台 UI
样式	无描边或细描边，保证移动端可读

字幕不是逐字稿。删掉“嗯、那个、其实”，观众看得更顺。

三、音乐与音量

- 人声为主时，音乐压到 −18 dB 左右；
- 无人声段落，音乐可抬到 −12 dB；
- 转场处用 0.3–0.5 秒淡入淡出；
- 不要在整条片子里一直堆音乐，留白更有力。

一个简单检查：手机外放，能否听清每个字。听不清，就是人声不够突出。

四、剪辑顺序

1. **粗剪**：按分镜排好，只看故事是否成立；
2. **精剪**：修节奏，删掉所有“可删”的帧；
3. **声音**：配音、音乐、音效；
4. **字幕**：对齐语义，不追着字走；
5. **调色**：统一色温与对比；
6. **导出**。

顺序不要乱。先解决故事，再解决好看。

五、转场怎么用

- **硬切**：90% 的场合都够用；
- **动作接**：上一镜动作未完，下一镜继续；
- **声音先入**：下一镜的声音提前 0.5 秒；

- **花式转场**：只在风格需要时用，不要每镜都加。

转场是标点，不是装饰。

六、导出参数

平台	比例	分辨率	帧率
竖屏短视频	9:16	1080×1920	30fps
横屏视频	16:9	1920×1080	30/60fps
方图	1:1	1080×1080	30fps

码率按平台推荐略高一点，给二次压缩留余量。导出前检查：首帧、尾帧、字幕安全区、音量峰值。

七、30 分钟练习

1. 写 60 字口播稿并生成配音（10 分钟）；
2. 按语义切成 4–5 屏字幕（8 分钟）；
3. 把 6 个镜头粗剪成 15 秒（8 分钟）；
4. 加音乐、压音量、导出竖屏版（4 分钟）。

八、检查清单

- [ ] 配音没有吞字与断句错；
- [ ] 字幕每行不超过 16 字；
- [ ] 人声在手机外放可听清；
- [ ] 转场以硬切为主；
- [ ] 导出比例与平台一致；
- [ ] 首尾帧完整，无黑屏。

九、延伸阅读

- 知识库 1.5 AI 视频

- 知识库 1.7 AI 数字人+语音合成
- 知识库 1.6 AI 音乐

> 下一讲：文生 3D 与资产准备。你会知道生成、扫描、建模三条路各自适合什么。
