KECUDO

PREMIUM

III

精品版

AI 电影的声音、音乐与配音

AI 电影制作 · 第三课



用 AI 工具做对白、音效、音乐、混音与配音——以及让电影原声可以安心发布的权利规则。

2026 版 · 创作者系列

本课内容

目录

第一部分 · 基础

1. 为什么声音是画面的另一半 4
2. AI 音频工具箱 6
3. 声音规划：对位会议与声音手册 8

第二部分 · 核心技能

4. 对白：录制、清理与 AI 修复 10
5. 用 AI 做音效与拟音 12
6. 环境声：分层搭建一场风暴 14
7. 音乐（上）：用 AI 工具配乐 16
8. 音乐（下）：作曲家与授权 18
9. 混音：电平、空间与响度 20

第三部分 · 进阶

10. 用 AI 声音做译制与本地化 22
11. 字幕、听障字幕与无障碍 24
12. 权利、同意与 AI 音频法规 26
13. 案例：完整的声音制作 28

自我测评

A 部分 · 基础 30
B 部分 · 计算 31
C 部分 · 像声音设计师一样思考 32
D 部分 · 真实工具与规则 33
E 部分 · 找出问题 34
你的评级 35

● ○ ○ 基础

第 1 章

为什么声音是画面的另一半

观众对模糊的画面、奇怪的一帧或一只怪异的手，远比对糟糕的声音宽容。含糊的对白、一阵嘶嘶声、每次剪接都变调的风暴，会立刻让人觉得业余。在 AI 电影里问题更突出：生成的片段自带音频，而每段听起来都略有不同。

声音能做到、画面做不到的事

- **告诉你该有什么感受。**同一个漆黑楼梯的镜头，配上噼啪的炉火很温馨，配上低沉的嗡鸣就很恐怖。
- **把镜头连在一起。**一条连贯的环境声，让来自三个模型的十个片段像同一个地方。
- **把世界延伸到画外。**海鸥、远处的雾笛、楼下的一扇门——观众会想象看不见的东西。
- **引导注意力。**一个声音提示会让我们在画面要求之前就转头去看。
- **承载时间。**钟声、呼吸、脚步决定了一场戏的节奏。

原声的五个组成部分

部分	包含	AI 能帮上的
对白	台词、呼吸、反应、旁白	清理、人声分离、临时配音、经同意的补录
音效	硬音效（雷声、门）与声音设计	文字生成音效、变体、检索
拟音	画面中的脚步、衣物、手拿道具	生成拟音、按画面对时
环境声	房间底噪、天气、地点感	长段底声、无缝循环
音乐	配乐与歌曲	临时配乐、小样、获授权的 AI 音乐

> **一句话总结：**把所有生成片段的音频静音，有意识地重建原声。片段自带的声音只是草稿，不是原声。

自己测一测：静音测试

取你电影里任意 30 秒，用三种方式看：

1. **完全静音。**注意它显得多平、多慢。
2. **用生成片段自带的音频。**注意每个剪接点上音色、响度和「空间感」的跳变。
3. **所有镜头下铺一条连贯的环境声底。**注意剪接点如何消失。

大多数人会发现，在加任何音乐或音效之前，第 3 步就已经让段落脱胎换骨。这就是有意设计的原声的力量。

声音在 AI 制作流程中的位置

阶段（第一课）	声音工作
剧本	把声音写进剧本；标出画外声
预演 / 动态分镜	用于定时的临时配音、临时音乐、关键音效
拍摄与生成	干净的现场对白、房间底噪、补录台词
剪辑	对白剪辑、临时混音、对位会议
声音后期	音效、拟音、环境声、配乐、终混
交付	分轨、不同响度版本、字幕、译制

KECUDO 影业 · 第一次试听

《守灯人之光》锁定画面的版本用的是每个片段自带的音频。风暴段落里，风声在 60 秒内变了 14 次；一个 AI 镜头在飓风中出现了海鸥；玛伦的实拍对白很干净，但听起来像在小房间里，而她明明应该在石塔中。团队把所有片段音频静音，从零开始做声音——这就是本书的内容。

常见的 AI 音频破绽

- 环境声在每次剪接时跳变。
- 不可能的声音（风暴中的鸟叫、海岛上的车流）。
- 人声音色在镜头之间变化。
- 音效早几帧或晚几帧。
- 音乐「泛泛」，不跟着故事走。
- 一切同样响——没有起伏，没有寂静。

> **寂静也是一种声音。**原声里最有力量的时刻，往往是一切都停下来的时候。规划好你的寂静放在哪里。

● ○ ○ 基础

第 2 章

AI 音频工具箱

AI 音频工具可以分为六类工作。多数制作每类都需要一个工具，再配上常规的剪辑与混音软件。和视频一样，先想「工作」，再想「品牌」。

工作	工具做什么	典型电影用途
语音生成	文字转语音、声音设计	临时对白、旁白草稿、动态分镜
变声	把一个声音转换成另一个（语音到语音）	经同意的补录、角色声音
清理与分离	去噪、去混响、去音乐；拆分声部	拯救现场对白
音效	文字生成音效；可检索的音效库	硬音效、拟音、环境声底
音乐生成	提示词 → 完整曲目或分轨	临时配乐、小样、获授权的配乐
译制与转写	语音转文字、翻译、配音	字幕、外语版本

一体化平台

不少平台如今把多项工作打包在一起。例如 ElevenLabs 已从文字转语音发展为集声音克隆、译制、转写、音乐、音效和完整音频编辑器于一体的平台。打包很方便，但共享额度池会让月度成本更难预测——留意各项工作的用量。

资料来源：Unite.AI 对 ElevenLabs 的评测（2026）。

选择音频工具

- **质量：**在好耳机上听也真实吗，而不只是在笔记本扬声器上？
- **控制：**能设定情绪、时长、长度、分轨、循环吗？
- **输出格式：**48 kHz WAV（而不只是压缩的 MP3）？
- **权利：**你的套餐允许商用吗？输出归谁？
- **声音保护：**克隆声音前会验证同意吗？

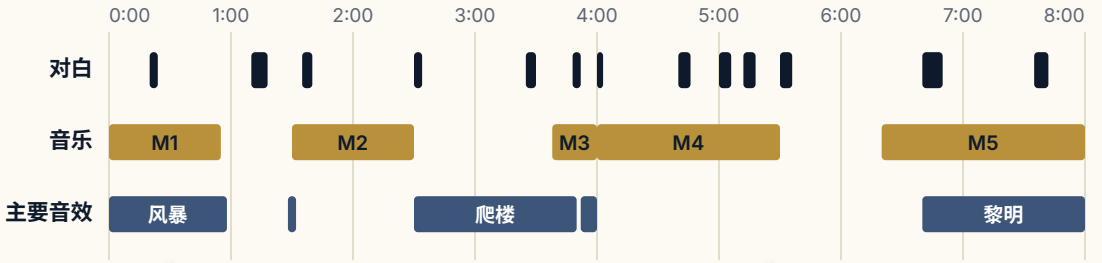
● ○ ○ 基础

第 3 章

声音规划：对位会议与声音手册

对位会议（spotting session）是导演、剪辑师和声音团队一起看锁定的剪辑版，精确决定每段音乐、每个关键音效和每段寂静放在哪里。它把「加点声音」变成一份可以排期、可以预算的计划。

FIG 3.1 · 《守灯人之光》（8:00）的对位图



段落	名称	起止	时长	来源
M1	风暴	0:00–0:55	55 秒	AI 弦乐铺底 + 作曲家大提琴
M2	独自	1:30–2:30	60 秒	八音盒动机 + AI 铺底
M3	重燃	3:38–4:00	22 秒	作曲家：独奏大提琴
M4	回忆	4:00–5:30	90 秒	八音盒动机 + AI 弦乐
M5	黎明	6:20–8:00	100 秒	作曲家大提琴 + AI 弦乐
合计	**5 段**		**327 秒**	**占 480 秒的 68.1%**

虚构示例——Kecudo 影业。

> **注意那些空白。**音乐覆盖了全片的 68.1%。空白处——0:55 空荡的厨房、爬楼段落的前半段、黎明的海岸——和音乐段落一样是刻意为之。

●●○ 核心

第 5 章

用 AI 做音效与拟音

文字生成音效工具能在几秒内把一句描述变成声音：「石塔里一扇沉重的铁门吱呀打开，外面刮着风」。结合一个已授权的音效库和几段自己的录音，一个小团队就能做出过去需要整个声音部门才能完成的音效。

写一条音效提示词

部分	问自己	示例
声源	什么在发声？	沉重的铁门
动作	发生了什么？	慢慢吱呀打开，然后撞上石墙
空间	在哪？多大？	在狭窄的石塔内
距离	近还是远？	近处，略偏左
特征	质地、年代、状态	生锈、老旧、沉重
长度	几秒？	4 秒

一扇生锈的沉重铁门慢慢吱呀打开，然后撞上石墙；在狭窄的石塔内；近处，略偏左；外面刮着风；干声，无音乐；4 秒。

硬音效与拟音

- **硬音效**是大而具体的声音（雷声、摔门、灯点燃）。AI 和音效库都是很好的来源。
- **拟音**是演员的细小人声动作：脚步、衣物、手碰道具。它必须逐帧匹配画面上的动作。

> **生成变体。**同一个脚步声绝不能连用两次——耳朵会听出重复。每个重复出现的声音生成 6–10 个变体并轮换使用。

自我测评

A 部分 · 基础

A1. 什么是「对位会议」（spotting session）？

A. 一次调色审看

B. 观看锁定剪辑版，决定每段音乐、关键音效和寂静放在哪里

C. 一次配音选角

D. 一次响度测试

A2. 什么是拟音？

A. 背景音乐

B. 雷声和爆炸

C. 与画面动作同步的细小人声动作：脚步、衣物、道具

D. 房间底噪

A3. D/M/E 分轨是什么？

A. 分开交付的对白、音乐和音效轨

B. 三种话筒

C. 草稿、混音和导出版本

D. 译制、母带和编码

A4. LUFS 衡量的是什么？

A. 帧率

B. 文件大小

C. 音高

D. 感知响度，在整个节目上取平均

A5. 隐藏式字幕（CC）与普通字幕的区别在于，它还包含……

A. 只有翻译

B. 说话人标识和声音描述，如［雷声］

C. 导演评论

D. 章节标题

K

KECUDO

感谢你与 Kecudo 一起学习。

系列下一课 · 第四课

AI 电影的发行与变现：电影节、平台、观众、为你的 AI 服务定价，
以及把一部短片变成一份事业。

KECUDO.COM