

免费资源

配乐之前：为 AI 电影执导声音

我的声音工作指南：在为场景加入配乐之前，先处理原生音频、声音远近与有意设计的静默。

Jeff Zambrano.

AI 电影导演 · jeffzambrano.com

适用读者

使用 AI 图像和视频模型创作电影、剧集及品牌影片的导演与创作者。

解决的问题

画面看似已经完成，生成的声音却缺乏层次，配乐只是掩盖问题，并没有让动作更清楚。

包含内容

- 原生音频清单
- 声音主次
- 角色与材质的声音词汇
- 静默设计
- 人声的空间位置
- 剪辑点之间的环境声衔接
- 聆听测试
- 可重复使用的场景工作表

目录

01 在音乐进入之前，先听听场景本身

02 先确认模型实际交付了什么

03 给每个声音明确的位置和任务

04 让身体和材质讲故事

05 设计静默，而不是留下空洞

06 让人声属于这个镜头

07 让空间感跨过剪辑点

08 使用具体的声音指令

09 判断音乐究竟能带来什么

10 填写场景工作表

01 在音乐进入之前，先听听场景本身

当一个场景显得还没完成时，加入音乐往往是最容易想到的下一步。但这样也可能让真正的问题更难被听见：呼吸没有空间，物体的声音与它的重量脱节，或背景声每到一个剪辑点就变了。配乐可以覆盖这一切，场景却依然难以读懂。

我首先决定观众需要听见什么。在我为《Borrowed Wounds》制定的早期声音方案中，呼吸、衣料、远处的钢铁声和风声各自承担明确的戏剧任务。方案主张不用配乐，因为这些声音已经足以传达动作。成片的演职员表中列有原创音乐；这里借鉴的是早期的一次声音规划，并非在描述最终的声音制作。这里沿用的是这套规划方法，并不把方案中拟定的时间当作成片的实测数据。

这份指南适用于画面已能提供具体依据的场景。完成后，你会明确哪些声音留下，哪些需要改变，哪些还要试听。它是一种执导声音的方法，并不承诺模型会执行每一条声音指令。

02 先确认模型实际交付了什么

我保留原始片段，在加入任何内容之前，先听它的原生音频。首先要判断的不是声音是否出彩，而是里面究竟有什么：脚步、衣料摩擦、呼吸、空气底声、人声，或一层不需要的音乐。

把每个听得见的元素分别记下来。如果两种声音混在一起，也要记录。能在嘈杂背景中听见人声，不代表手上就有可独立编辑的干净人声。方案必须如实描述现有素材，也包括它的限制。

接着，我把结果与该镜头要求的声音作比较。指令只能证明创作意图，不能证明结果已经实现。我会记录缺失、意外出现的声音，以及它们出现时机的差异，也会比较前后镜头的背景嘶声。一个片段单独听可能没问题，一接上相邻镜头，录音质感的差异就显露出来。

- [ ] 我已在没有额外声音层的情况下听过原生音轨。
- [ ] 清单已区分独立元素和混在一起的声音。
- [ ] 我已记录要求与实际音频之间的差异。
- [ ] 原始素材仍保留，可随时比较。

03 给每个声音明确的位置和任务

对于准备保留的每个元素，我都会写下它的相对音量、立体声宽度和听感距离。这是三个不同的决定。声音可以很轻却很近，也可以很突出，但仍明显来自远处。把所有声音调大，并不能建立听觉视点。

我用直白的描述：前景或背景、居中或铺开、近处或远处。这些是在场景内部建立的关系，不是通用的混音参数。我不会仅仅为了让工作表显得更专业，就填入一个分贝数值。

然后，我确定每段动作由哪个声音引导观众。在原始声音方案中，可以先让外部事件占据主导，等身体状态发生变化，再交给呼吸。关键是这次交接。如果所有声音始终同样突出，交接就会被掩盖。

方案将短镜头的注意力集中在两三个主要元素上。我把它当作取舍的约束，而不是禁止背景中存在其他声音。如果还说不清哪个元素占主导，这一段的声音执导就还没完成。

- [ ] 每个保留的元素都有明确的音量、宽度和距离。
- [ ] 我能指出镜头各段的主导声音。
- [ ] 声音重点的转移，对应动作中可以辨认的变化。

04 让身体和材质讲故事

我从场景中实际存在的事物建立声音词汇。衣料需要明确材质和状态，脚步需要明确地面，呼吸需要明确身体状况。这些区分比一句“营造戏剧氛围”更能指导工作。

声音方案区分干燥、硬挺的亚麻与状态已经改变的布料，也区分松散沙地和夯实泥地上的移动。目的在于让观众读懂这个世界，而不是为每个看得见的物体收集音效。一个声音值得存在，是因为它所表现的接触、运动或变化有意义。

对于话不多的角色，我会比较戏剧变化前后身体发出的声音。绵长的鼻息与张嘴急促呼吸，表达的是不同的身体状态。我会决定变化发生在哪里，后一个镜头是否延续了它。不能因为新片段开始，就把角色的身体状态重置。

处理发声反应时，我也保持同样的克制。一次呼吸或用力声，需要对应动作和意图，不能变成贴在每次运动上的噪声。如果嘴部可见，声音表演就必须与面部和身体的表现一致。

- [ ] 每个突出的接触声都有可辨认的材质或表面。
- [ ] 呼吸或声音质感的变化，对应场景中的变化。
- [ ] 后一个镜头延续了前一个镜头建立的身体状态。

05 设计静默，而不是留下空洞

写下“静默”时，我会说明什么停止，什么保留。远处的钢铁声停止、风声继续，与所有声音一起消失，是两件不同的事。前者可以在同一空间内转移注意力，后者连空间感也一并拿走。我会有意识地作出选择。

我还会说明声音如何停止。突然中断与渐弱消失，讲述的是不同的事情。如果动作需要戛然而止，平顺的淡出反而可能削弱我希望观众注意的事件。

我根据动作、信息揭示或台词来安排安静的间隙，再结合实际画面与声音判断时长。原始方案明确将部分拟定的间隙标为有待试听的设计决定。我保留这个区别：方案中写了一个时长，并不能证明停顿有效。

静默不一定要让人明显察觉它是一种效果。有时，只需让一个竞争注意力的声部退下，微小的声音就能被听清。我听的是它带来的结果：现在能否听见场景要我注意的东西？

- [ ] 我已列明停止和保留的声音层。

- [ ] 已明确过渡采用突然中断还是淡出。
- [ ] 安静的间隙服务于具体事件，并已结合该事件测试。

06 让人声属于这个镜头

清楚的人声也可能像贴在画面上的。我会将它的听感距离和混响，与画面中的空间比较。近距离录音并不一定适合站在开阔场地远处的说话者。

提高人声音量之前，我先决定哪些声音要让位。原始方案在台词周围留出空间，而不是让对白必须压过持续不断的背景层。这样，台词就成为场景注意力变化的一部分。

我把台词的表达方式与声学位置分开描述。平淡、轻声地说一句话，是表演决定；听起来近还是远、在室内还是室外，是空间决定。两者都要对照生成镜头检查，因为实际构图可能与计划不同。

不能因为提出了要求，就认定人声或音效生成器一定能准确表现距离和混响。如果结果不属于画面中的空间，那仍是需要混音处理的问题。一条写得有说服力的指令，不等于已经完成声学匹配。

- [ ] 人声的距离和混响符合画面中的空间。
- [ ] 其他声音为台词留出了位置。
- [ ] 我已对照实际镜头检查交付的声音表演。

07 让空间感跨过剪辑点

分别生成的片段可能带来各不相同的背景质感。在认为段落已经完成之前，我会听它们的衔接。即使画面连续，嘶声或环境声的突变也可能暴露剪辑点。

原始方案是在检查原生音频之后，才建立共同的环境底声。这个顺序很重要：先听素材，才能决定如何连接。我会与背景较干净的镜头比较，处理不一致的噪声，而不是先认定所有原生声音都必须丢弃。

每个衔接处，我都写明什么延续、什么变化，以及原因。地面或地点变化，可以成为声音变化的理由；仅仅换了一个生成片段，不可以。观众应该听见故事中的空间关系和动作，而不是片段的边界。

然后，我闭上眼听完整个组接后的场景。如果背景在没有叙事理由的情况下跳变，让我能听出剪辑点，我就回到那里检查。这是连续性测试，并不要求掩盖每一次有意安排的声音转换。

- [ ] 背景变化源于场景，而不是生成片段的边界。
- [ ] 持续的声音跨过剪辑点后仍保持一致。
- [ ] 我已不看画面，听完整个段落。

08 使用具体的声音指令

我希望音效指令描述可听见的事件：声源、动作、质感和听觉位置。下面的例子来自声音方案，要求的是布料接触声，而不是某种情绪：

```
An open dry palm dragging slowly and continuously across a stiff, coarse, heavily textured linen surface, a low steady friction sound with no wetness, no music.
```

我只在确实符合该材质和动作的地方使用它。这是原始方案中的示例，不是经验证的特定生成器输出保证。生成之后，仍要对照画面检查接触、时机和质感。再动听的音效，只要属于另一种材质，就是错误的音效。

这条指令排除音乐，是为了单独获取所需音效，并不决定最终场景是否应该配乐。配乐的取舍，要从整个场景判断。

09 判断音乐究竟能带来什么

当动作、听觉位置和连续性已经清楚，我才考虑音乐。除了重复身体、人声和环境已经传达的情绪，它还能贡献什么？此时我能更有依据地决定，因为已经听见了场景独立成立时的表现。

《Borrowed Wounds》的声音方案反对配乐，有具体理由：主要声音已占据听众的注意力，静默承担戏剧分量，音乐只会重述信息。我不会把这个选择变成所有 AI 电影都必须不用配乐的规则。

我把画面和声音分开测试。隐藏画面，听声音能传达哪些动作、空间和身体状态的信息；关掉声音，看哪些信息随之消失。然后恢复两者，判断它们的关系。如果一个声音层没有带来有用信息，我会重新考虑它的作用，而不是仅因制作精致就留下。

- [ ] 我能说明拟加入的配乐为这个场景贡献什么。
- [ ] 我已分别测试画面和声音，再将两者一起检查。
- [ ] 配乐决定依据场景的需要作出。

10 填写场景工作表

剪辑时，我把这张表放在手边，根据实际观看和聆听填写。素材尚未出现时，未知项就明确留着。它记录决定，但不把方案当作已获认可的结果。

决定事项	我的记录
观众必须感知的戏剧变化	
实际听见的原生声音	
原始音频中混在一起的元素	
要保留、替换或添加的声音及理由	
每个保留声音的相对音量、宽度和距离	
变化前后的主导声音	
角色状态与材质状况	
什么停止、什么保留，以及静默如何开始	
人声的表达方式、距离和混响	
延续到下一个剪辑点之后的声音	
音乐能够贡献什么	
已完成的聆听测试及尚未解决的决定	

只有表中所写与实际听见的内容一致，我才结束记录。如果片段尚未生成、混音尚未组接，或过渡尚未测试，相应决定就继续保留为未定。真正有用的成果，是场景中的每个声音都有存在的理由，同时清楚标出剩余工作。

JEFF ZAMBRANO. · AI 电影导演

一起完成你的电影

我用 AI 执导电影、剧集和品牌作品，从剧本到最终剪辑。
告诉我你的想法，我会亲自回复。



委托制作电影、剧集或预告片
请先说明影片形式、预计时长和交付日期，再分享已有的故事或素材。

资助 Borrowed Wounds 长片
请告诉我你考虑投资、联合制作还是赞助这部 90 分钟的电影，我会把逐镜头核算的制作计划发给你。

品牌影片
请介绍品牌与影片的目的，并在已确定的情况下，说明目标观众及发布日期。

AI 平台或赞助
请介绍你的工具或模型，再简述你设想的影片或赞助合作。

电影节、媒体或发行
请介绍你的活动、媒体或发行机构，并注明日期，以及提议所涉及的影片或素材。

hello@jeffzambrano.es
WhatsApp +34 688 351 935
jeffzambrano.com