

FREE RESOURCE

Before the Score: Directing Sound in an AI Film

My working guide to native audio, sound perspective and deliberate silence before adding music to a scene.

Jeff Zambrano.

AI film director · jeffzambrano.com

WHO IT IS FOR

Directors and creators making films, series and brand pieces with AI image and video models.

THE PROBLEM IT SOLVES

A scene looks finished, but its generated audio feels flat and music covers the problem instead of clarifying the action.

WHAT YOU GET

- Native audio inventory
- sound hierarchy
- character and material vocabulary
- deliberate silence
- voice perspective
- ambience across cuts
- listening tests
- reusable scene worksheet

Contents

01 Give the scene a chance before the music arrives

02 Start with what the model actually delivered

03 Give each sound a position and a job

04 Let the body and the material tell the story

05 Design the silence instead of leaving a hole

06 Make a voice belong to the shot

07 Carry the place across the cut

08 Use a concrete sound instruction

09 Decide what music would contribute

10 Fill in the scene worksheet

01 Give the scene a chance before the music arrives

When a scene feels unfinished, music is an easy next move. It can also make the real problem harder to hear. A breath has no space, an object sounds detached from its weight, or the background changes at every cut. A score can sit over all of that while the scene remains difficult to read.

I start by deciding what the audience needs to hear. In an early sound plan for Borrowed Wounds, breath, cloth, distant steel and wind carry specific dramatic jobs. The plan argues for leaving music out because those sounds already carry the action. The finished film credits original music; I am drawing on an earlier planning exercise here, not describing the final soundtrack. I use that planning method without presenting its proposed timings as measurements of the finished film.

This guide is for a scene whose images already give you something concrete to work against. You will leave with decisions about what stays, what changes and what needs listening tests. It is a method for directing a soundtrack, not a promise that a model will obey every sound instruction.

02 Start with what the model actually delivered

I keep the original clip and listen to its native audio before adding anything. The first question is not whether the soundtrack is impressive. I want to know what is in it: a footstep, a fabric movement, a breath, background air, a voice or an unwanted musical layer.

Write down each audible element separately. If two sounds are fused together, record that fact. Hearing a voice inside a noisy background does not mean you have an isolated voice available to edit. My plan must describe the material I have, including its limitations.

Then compare the result with the sound requested for the shot. An instruction is evidence of intention, not evidence of delivery. I note missing sounds, unexpected sounds and any difference in when they occur. I also describe the background hiss relative to the surrounding shots. A clip can sound acceptable alone and reveal a different recording texture as soon as it meets its neighbour.

- [ ] I have listened to the native track without added layers.
- [ ] My inventory distinguishes separate elements from sounds fused together.
- [ ] I have recorded differences between the requested and delivered audio.
- [ ] The original remains available for comparison.

03 Give each sound a position and a job

For every element I intend to keep, I write its relative level, stereo width and apparent distance. These are different decisions. A sound can be quiet and close, or prominent while still belonging to an event farther away. Turning everything up does not establish a point of view.

I use plain descriptions: foreground or background, centred or spread, near or distant. These are working relationships within the scene, not universal mix settings. I do not assign a decibel value simply because a worksheet looks more technical with numbers in it.

Next I identify what leads the listener through each stretch of action. In the source sound plan, an external event can lead first and breathing can take over when the body's condition changes. The important decision is the handover. Keeping everything equally prominent would conceal it.

The plan concentrates attention on two or three leading elements in a short shot. I treat that as a discipline for choosing, not a ban on a background that contains other sounds. If I cannot say which element leads, I have not finished directing the passage.

- [ ] Every retained element has a stated level, width and distance.
- [ ] I can name the leading sound in each stretch of the shot.
- [ ] A change of emphasis follows an identifiable change in the action.

04 Let the body and the material tell the story

I build a sound vocabulary from the things that actually exist in the scene. Cloth needs a material and a condition. Footsteps need a surface. Breathing needs a physical state. Those distinctions give me more useful instructions than asking for a dramatic atmosphere.

The sound plan distinguishes dry, stiff linen from fabric whose condition has changed. It also distinguishes movement on loose sand from movement on packed earth. The purpose is to make the world legible, not to collect effects for every visible object. A sound belongs because its contact, movement or change matters.

For a character who speaks little, I compare the body's sound before and after the dramatic change. Long nasal breathing and short open-mouth breathing describe different physical states. I decide where that change occurs and whether the following shot preserves it. I do not reset the character's body just because a new clip begins.

I use the same restraint with vocal reactions. A breath or an effort sound needs a gesture and an intention. It should not become a noise pasted onto every movement. If the mouth is visible, the performance has to belong to what the face and body are doing.

- [ ] Every featured contact sound has an identifiable material or surface.
- [ ] Changes in breathing or texture correspond to changes in the scene.
- [ ] The next shot preserves the physical state established by the previous one.

05 Design the silence instead of leaving a hole

When I write silence, I specify what stops and what remains. Distant steel stopping while wind continues is a different event from every layer disappearing. The first can redirect attention within a place; the second removes the place as well. I choose between them deliberately.

I also specify how a sound stops. An abrupt interruption and a fade tell different stories. If the action calls for a hard stop, a graceful fade can soften the very event I need the audience to notice.

I place the quiet interval in relation to an action, a revelation or a line. I then judge its duration against the actual picture and sound. The source plan explicitly identifies some proposed gaps as design decisions awaiting a listening test. I keep that distinction. A duration written in a plan is not proof that the pause works.

Silence does not have to announce itself as an effect. Sometimes the useful change is simply that a competing layer gives way and a small sound becomes readable. I listen for the consequence: can I now hear the thing the scene asks me to notice?

- [ ] I have named the layers that stop and those that remain.
- [ ] The transition is identified as an interruption or a fade.
- [ ] The quiet interval supports a specific event and has been tested against it.

06 Make a voice belong to the shot

A clear voice can still sound pasted onto an image. I compare its apparent distance and reverberation with the space on screen. An intimate recording does not automatically fit a speaker standing farther away in an open location.

Before raising the voice, I decide what yields to it. The source plan opens space around spoken lines instead of treating dialogue as a layer that must overpower a continuous bed. That makes the line part of the scene's changing attention.

I describe the delivery separately from its acoustic placement. A flat, quiet line is a performance decision. Whether it sounds close, distant, enclosed or outdoors is a spatial decision. Both need checking against the generated shot, which may differ from the planned framing.

I cannot assume a voice or effects generator will reproduce distance and reverberation faithfully because I requested them. If the result does not belong to the picture, that remains a mixing task. A convincing instruction is not a finished acoustic match.

- [ ] The voice's distance and reverberation agree with the depicted space.
- [ ] Competing sounds leave room for the line.
- [ ] I have checked the delivered performance against the actual shot.

07 Carry the place across the cut

Separately generated clips can bring separate background textures. I listen across their joins before treating the sequence as finished. A sudden change of hiss or ambience can expose the edit even when the visual continuity holds.

My source plan builds a common ambient bed after inspecting the native audio. That order matters: I need to hear the material before deciding how to connect it. I compare the background with the cleaner shots and address inconsistent noise without assuming that every native sound must be discarded.

For each join, I write what continues, what changes and why. A change of surface or location can justify a change of sound. A new generation alone does not justify it. The listener should hear the story's geography and action, rather than the boundaries of the clips.

I then listen through the assembled scene with my eyes closed. If I can locate a cut because the background jumps without a story reason, I return to that join. This is a continuity test, not a demand to hide every intentional sonic transition.

- [ ] Background changes correspond to the scene rather than the generation boundaries.
- [ ] Continuing sounds keep their identity across the join.
- [ ] I have listened through the sequence without looking at the image.

08 Use a concrete sound instruction

I want an effects instruction to describe an audible event: source, action, texture and perspective. This example comes from the sound plan. It requests a fabric contact rather than a mood:

```
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.
```

I use it only where that material and action belong. It is a source example, not a verified guarantee of a particular generator's output. After generation, I still compare the contact, timing and texture with the picture. A beautiful effect for a different material is the wrong effect.

The absence of music in this instruction isolates the requested effect. It does not decide whether the finished scene should have a score. That decision belongs to the scene as a whole.

09 Decide what music would contribute

Once the action, perspective and continuity are readable, I consider music. I ask what it would contribute beyond repeating the emotion already carried by the body, voice and environment. I can make a stronger decision now because I can hear what the scene does on its own.

In the Borrowed Wounds sound plan, the argument against a score is specific: the leading sounds already occupy the listener's attention, the silence carries dramatic weight, and music would restate information. I do not turn that choice into a rule that every AI film must be unscored.

I test picture and sound separately. With the picture hidden, I listen for information about action, space and physical state. With sound muted, I watch what information disappears. Then I restore both and judge their relationship. If a layer adds no useful information, I reconsider its role before keeping it merely because it sounds polished.

- [ ] I can state what a proposed score contributes to this scene.
- [ ] I have tested picture and sound separately, then together.
- [ ] The music decision follows the scene's needs.

10 Fill in the scene worksheet

I keep this sheet beside the edit. I fill it from listening and viewing, leaving unknowns explicit until the material exists. It records decisions without pretending that a plan is an approved result.

Decision	My entry
Dramatic change the audience must perceive	
Native sounds actually heard	
Elements fused together in the source audio	
Sounds to retain, replace or add, with reasons	
Relative level, width and distance of each retained sound	
Leading sound before and after the change	
Character state and material conditions	
What stops, what remains and how the silence begins	
Voice delivery, distance and reverberation	
Sound that crosses the next cut	
What music would contribute	
Listening tests completed and unresolved decisions	

I close the sheet only when its claims match what I have actually heard. If the clip has not been generated, the mix has not been assembled or the transition has not been tested, I leave that decision open. The useful result is a scene whose sound has a reason to be there, with the remaining work plainly identified.

JEFF ZAMBRANO. · AI FILM DIRECTOR

Let's make your film

I direct films, series and brand pieces made with AI, from the script to the final cut. Tell me what you have in mind and I will answer myself.



Commission a film, series or trailer

Start with your film's format, intended length and delivery date, then share the story or material you already have.

Fund Borrowed Wounds, the feature

Tell me whether you are thinking of investing, co-producing or sponsoring the 90-minute film, and I will send you the production plan costed shot by shot.

A film for a brand

Introduce the brand and the purpose of its film, including the intended audience and release date where known.

AI platform or sponsorship

Introduce your tool or model, then outline the film or sponsorship you have in mind.

Festival, press or distribution

Introduce your event, publication or distributor, specifying the date and the film or material your proposal involves.

hello@jeffzambrano.es

WhatsApp +34 688 351 935

jeffzambrano.com