PROMPT CRAFT

Audio prompts

Tempo, instrumentation and the structure a generated cue needs before it can be cut against picture.

HOW TO USE THIS

Five habits that do most of the work

- Work down the aspects in order. The order is the recommended order in the prompt too.
- One clause per aspect. A prompt is a list of decisions, not a paragraph of atmosphere.
- Prefer a number over an adjective wherever a number exists: 116 BPM over upbeat, 85mm over flattering, 15 seconds over short.
- Put the negatives last, in one sentence beginning with No.
- Reuse the identical phrasing across a batch. Paraphrasing is what makes a set of images look unrelated.

THE ORDER

Work down these, in this order

The order below is also the order that works inside the prompt itself. Each one is a decision; state it in one clause.

- Duration, Role, Mood, Key and tonality
- Tempo, Instrumentation, Arrangement and dynamics, Vocals
- Ending

ASPECT 1 OF 11

Duration

A generator will otherwise give you a loop of arbitrary length that has to be cut, which is where the musical ending gets lost.

Vocabulary	15 seconds, 30 seconds, 60 seconds, 90 seconds, 120 seconds
Phrases that work	90-second bed.
	30 seconds with a clean ending.
	15-second sting.

ASPECT 2 OF 11

Role

A bed under narration and a standalone track need opposite decisions about melody and midrange, so state the job before the style.

Phrases that work

Instrumental music bed to sit under narration.

Standalone track, no voice over it.

Short sting to close a scene.

ASPECT 3 OF 11

Mood

Pairs with tempo and key. Together those three settle most of the composition.

Vocabulary	warm, playful, cheerful, buoyant, wholesome, ambient, uplifting, driving, triumphant, minimal
Phrases that work	Cheerful, wholesome and playful.
	Calm and spacious.
	Suspenseful and restrained.

ASPECT 4 OF 11

Key and tonality

The cheapest emotional lever in audio. Major reads as safe and bright, minor as unresolved, and models respond to both.

Vocabulary	major key, minor key, modal, pentatonic
Phrases that work	Bright major key.
	Minor key, unresolved.
	Warm, clean major-key resolution in the final eight seconds.

ASPECT 5 OF 11

Tempo

A number is obeyed far more reliably than a word like upbeat.

Vocabulary	70 BPM, 90 BPM, 100 BPM, 116 BPM, 128 BPM
Phrases that work	Buoyant mid-tempo around 116 BPM.
	Slow, roughly 70 BPM.
	Driving, 128 BPM, steady four on the floor.

ASPECT 6 OF 11

Instrumentation

Listing instruments works better than naming a genre, which models interpret loosely.

Vocabulary	percussion, ukulele, piano, bass, handclaps, marimba, bells, strings, guitar, xylophone
Phrases that work	Marimba, ukulele, xylophone, light percussion and soft handclaps.
	Solo piano with a distant string pad.
	Deep electronic ambience with soft mechanical clicks.

ASPECT 7 OF 11

Arrangement and dynamics

Without this you get four bars looped flat. Naming a build and an ending is what makes it usable against picture.

Vocabulary	airy, midrange, headroom, no abrupt hits, narration-friendly, smooth transitions, gentle dynamics
Phrases that work	Airy and narration-friendly, gentle dynamics, no abrupt hits.
	Build subtle energy throughout, then a celebratory lift at the end.
	Clear midrange space left for a voice.

ASPECT 8 OF 11

Vocals

Generators add humming and wordless vocals unprompted.

Phrases that work

No vocals, no lyrics, no spoken words.

Wordless humming only, no intelligible words.

No prominent lead melody competing with the narration.

ASPECT 9 OF 11

Ending

Ask for a resolution.

Phrases that work

End on a clean major-key resolution, no fade.

Ring out and decay naturally.

Hard stop on the final beat.

ASPECT 10 OF 11

Voice and spoken line

Treat it as practice worth copying, not as a count.

Vocabulary	one voice per speaking entity, exact line, performance direction
Phrases that work	The gate is open. Move!
	same voice, calmer delivery
	Voice prompts contain the exact line; performance direction belongs in the model-supported prompt, not undocumented flags.

ASPECT 11 OF 11

Narration writing (spoken script)

Treat it as practice worth copying, not as a count.

Vocabulary	20-24 words, about 8-9 seconds, under roughly 9.5 seconds, plain spoken text only, no timecodes, emotion cues, parentheticals, or stage directions, Spell numbers out, hook through
Phrases that work	Write one plain line per clip, sized for about 8-9 seconds and normally 20-24 words. Keep every take under roughly 9.5 seconds. Use no timecodes, emotion cues, parentheticals, or stage directions. Spell numbers out.

AUDIO TAGS

ElevenLabs v3 audio tags

Eleven v3 text-to-speech, per ElevenLabs' own documentation

Voice and delivery	[laughs], [laughs harder], [starts laughing], [wheezing], [whispers], [sighs], [exhales], [sarcastic], [curious], [excited], [crying], [snorts], [mischievously]
Sound effects	[gunshot], [applause], [clapping], [explosion], [swallows], [gulps]
Unique and special	[strong X accent], [sings], [woo], [fart]

AUDIO TAGS

Suno structure tags

Suno lyric input. Widely used, and not published by Suno as far as we could verify. Not published by the vendor; convention.

Song sections	[Intro], [Verse], [Verse 1], [Verse 2], [Pre-Chorus], [Chorus], [Hook], [Bridge], [Outro], [End]
Instrumental and dynamics	[Instrumental], [Instrumental Break], [Break], [Interlude], [Drop], [Build-Up], [Breakdown]

NEGATIVES

Say what you want, not what you don't

Most media models expose no separate negative field, so a banned word is still a word in the prompt. Keep the list short and aimed at the failure you actually see.

- Text and branding: spoken words, words, other words, harsh textures, ad-libs that obscure the words
- Content and safety: scary sounds
- Audio: vocals, lyrics, sound effects, speech, music, spoken narration, voices, melody
- Other: medical sounds, abrupt peaks, dark, dramatic tension, heavy drums, long instrumental passages, sung parts, dialogue

Read the rest online

The web pages are the version that gets corrected. A sheet printed six months ago has not moved.

Cinematic language: lenses, film stocks, lighting and mood recipes.

Every measurement on the site carries the date it was read.

shotcraft.video/prompts/audio/