

One More Take

One actress. Every note. None of it filmed.

A page where you direct Mira, an invented actress, take after take. Claude Opus 5.5 cast her, shot every take and checked every line in three headless runs, driving Scenario's image, voice, music, lip-sync, 3D and video models over the Scenario MCP. Nobody real is in it: Mira, her cat and every voice were generated.

Every step, prompt, figure and timing in this document comes from the session's own notes, call-by-call recipe and Scenario job records.

72

takes on the page

16,972

CU in all

243

Scenario calls

69 / 69

spoken takes checked



As written, as your cat, as a marble bust, as anime, on a 1990s camcorder: five of the finished takes, frame 1.0 s.

One line. One whistle. Every note.

The concept was fixed before the session started: Mira, an invented woman in her early thirties, has one line, “Can I tell you a secret? I've never told anyone. I can't whistle.”, and tries to whistle at the end of every take. Every button on the page is a director's note: a language, a delivery, who she is, when it was shot, or a pair of notes at once.

HOW THE SESSION WORKED

- Everything invented: no real people, no real voices, no voice cloning of anyone real.
- Before each model, read its live schema and price the exact call with a free quote.
- Pick the best and newest model for each job, with Scenario's recommend and a search by newest.
- Look at every clip through sampled frames and transcribe every spoken take back to check the words.
- Measure the CU and seconds of every call.

Scenario was reached over the Scenario MCP: its tools for model search, recommendations, job checks and the project collection, and a small Python client built on the Scenario Blender Plugin's Scenario MCP connection, which quoted, submitted, waited on and logged every generation.

PILOT · 24 SEPTEMBER

Casting and the hardest cases: the hub take, her cat, an opera, Arabic, a whisper, a 1990s camcorder, a marble bust and five lip-sync models on the same audio.

53 calls · 3,089 CU · 42 min from first call to last clip

FULL SHOOT · 26 SEPTEMBER

Every deck from the pilot's measured routes: twelve languages, twelve deliveries, six who, six when, ten pairs, then a first working page.

133 calls · 6,524 CU · about 60 min of generation, three jobs in flight

PAIRS · 27 SEPTEMBER

Twenty more pairs, covering every combination of decks; ten of them run Seedance 2.5 on a finished take.

57 calls · 7,359 CU · about 70 min, three jobs in flight

A casting tape, not a stock photo

Two image models received the same prompt, a casting self-tape still lit by a window with a white bounce, which ends “the look of a real audition tape recorded in a small casting studio, not a stock photo, not a beauty ad.” Krea 2 Large's frame read as a real audition tape: the window, the reflector, unretouched skin. GPT Image 2.5 Sunburst's two were sharper and prettier and looked like stock photography. Krea won (12 CU, 54 s).



Chosen: Krea 2 Large. Not used: two GPT Image 2.5 Sunburst candidates.



The driver sheet, a GPT Image 2.5 Sunburst edit of the portrait.

GPT Image 2.5 Sunburst then edited the chosen portrait into a driver sheet: four views and three mouth shapes, closed, ah and whistle, with legible labels (13 CU). The sheet is the identity reference for every later image of her.

A portrait, a script, two prompts

P-Video Avatar turns one image and a script into a speaking take, with a preset voice (Sulafat) that became Mira's voice. The script ends in the whistle attempt, written as breath: "Fffffhhh... fwhhh...". A voice prompt directs the delivery and a video prompt directs the face.

Two seeds cost 17 CU each and about a minute each. Seed 9031 gave the clearer whistle pucker and became the hub; seed 4127 kept the same face and voice with a different performance. The same call with a whisper prompt produced a true, unvoiced whisper.

VOICE PROMPT, VERBATIM

Warm, intimate, slightly embarrassed, as if confiding in a close friend. Natural conversational pacing with short pauses between the three sentences. After the last sentence she purses her lips and tries to whistle, but only soft breathy air comes out, twice, then a tiny self-conscious laugh.

The ellipses made a 14 s take. For the full shoot the hub was re-cut to 10 s, the line and the first whistle puff, so the era edits fit and the five lip-sync comparisons could be reused; new takes used a tighter script.



The hub take, eight frames sampled by the session: speech, a laugh, the whistle pucker.

One face, four routes

LANGUAGE · 12

P-Video Avatar speaks British English, Spanish, German, Italian, Brazilian Portuguese, Japanese, Korean and Hindi from a language setting. For Arabic, Mandarin and Vietnamese, ElevenLabs Dubbing v2 dubs the hub audio and keeps her voice, then P-Video Avatar moves her mouth to it; French took that route after P-Video's French voice failed twice with server errors.

WHO · 6

GPT Image 2.5 Sunburst edits the portrait with the driver sheet into Mira at 85, as an oil painting, in claymation, as anime and as her cat; P-Video Avatar animates each one, stylised faces included. The marble bust took a 3D route. “At 8” was dropped: no clean child voice came out.

DELIVERY · 12

The voice prompt alone gives the whisper, laughing, furious, heartbroken, sarcastic, job-interview, sports commentator and bedtime takes. ElevenLabs Music v2.5 sings the opera and raps the rap, in another singer's voice, then P-Video Avatar sings along. Seed Audio 1.0, with the hub take as its voice reference, makes helium (pitch +12) and slow motion (speech rate -40).

WHEN · 6 AND PAIRS · 30

Seedance 2.5 re-renders the 10 s hub cut in six eras. A pair is two notes in one take, thirty across every combination of decks, from the same routes: the cat singing the opera, Mira at 85 rapping, an oil painting in a 1920s silent film.



Who source images, each an edit of the portrait with the driver sheet: her cat, at 85, oil painting, claymation, anime.

Does a mouth hold on stone?

1 CONCEPT

GPT Image 2.5 Sunburst carved her in white Carrara marble from the portrait and the driver sheet (13 CU).

2 MESH

Meshy 7.1 turned the concept into a textured 3D bust, 2K geometry with PBR maps (140 CU, 368 s). It made the round socle a square block.

3 ONE FRAME

Blender 5.2 imported the mesh with the Scenario Blender Plugin, stood it on a charcoal plinth under museum light and rendered one Cycles frame at 720 x 1280 (256 samples, 21.8 s).

4 LIP SYNC

Minimax H3 Max Lip Sync moved the stone to the hub audio (226 CU, 43 s): lips and jaw follow the words, the teeth come out as marble, bust and plinth stay rigid.

P-Video Avatar was the only model that acted the whistle on the bust, but the head swung and the plinth changed shape between frames, so it was rejected for rigid subjects. Sync-3 Avatar held like H3 at 395 CU and six and a half minutes.



Concept



Cycles still



The take

Change the decade, keep the mouth

Seedance 2.5 ran in reference mode: the 10 s hub cut as @video1, and a look key frame, a GPT Image 2.5 Sunburst edit of the hub's first frame (12 CU), as @image1. Each prompt lists what must stay, “every head movement, every blink and every mouth shape frame for frame”, then what must change: orthochromatic grain for the 1920s, tungsten skin and a JUN 14 1996 stamp for the camcorder. Each era cost 554 CU at 720p.



Look key frames: 1920s, 1950s, 1970s, 1990s, 2010s, 2026.

A FITTED RETIME

Every era take of the full shoot came back 3% fast, 233 frames for 240, and a 240/233 retime restored the mouth (0.98 on four eras, 0.92 and 0.95 on the webcam and vertical video). One pair later came back at real speed, so the retime is now fitted on the mouth itself.

PERIOD SOUND

A piano score from ElevenLabs Music v2.5 under intertitles for the silent 1920s; ElevenLabs Sound Effects 2 beds for newsreel crackle, TV hum, tape hiss and room tone; a beat and burned captions for 2026.

ONE STAMP FIX

On the Arabic 1990s stack, Seedance changed the stamp from 1996 to 1997 mid-take; the correct stamp was composited from the first frame. Gemini Omni 1.1 Flash Edit versions were made for comparison and not used.



Mouth checkpoints: the hub, the retimed Seedance 1990s take, the Gemini Omni comparison.

What failed, and what was done

Every clip was frame-checked, including a close sweep of the mouth in the last seconds, and every spoken take was transcribed back with Gemini 3.5 Transcribe: 69 of 69 say the right words, and the dubbed takes reword the line with the right meaning. Audio was mastered to -16 LUFS.

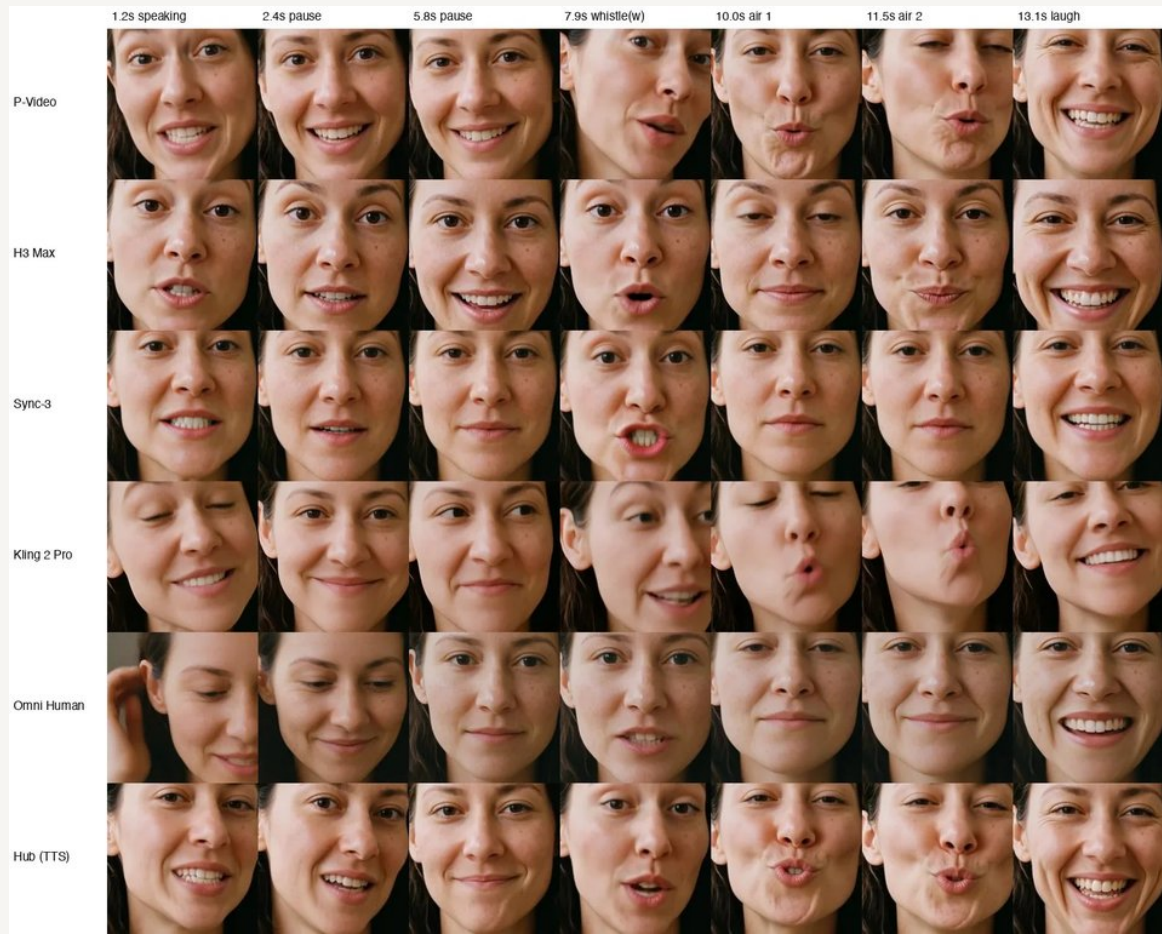
TAKE	WHAT WENT WRONG	WHAT WAS DONE
Korean, anime + Korean	Garbled Korean subtitles burned into the picture	Redone with a negative prompt against on-screen text
Claymation	The first image read as a real woman with sculpted hair	Image redone with a stronger puppet prompt
Spanish + commentator	A pasted, lighter mouth patch during the whistle	Redone with a new seed
French, Japanese whisper	Server errors twice on the same input	Route switched to ElevenLabs Dubbing v2
Italian + helium	Seed Audio at +12 garbled the first phrase, twice	The Italian take's audio raised an octave in ffmpeg
French + 1920s	Seedance turned her irises glassy silver	Take 2 with an explicit clause about her eyes
Rap + 2010s webcam	The mouth drifted and Seedance invented a whistle	Take 2 without the whistle, retime fitted on the mouth
Her cat (pilot)	Turned into a different cat in the silent tail	Cut where the sound ends
At 8	Seed Audio's child voice came back adult-pitched and quiet	Dropped from the Who deck



Claymation: the rejected first image and the redo.

Same face, same audio, five models

The page's last section plays one portrait and one audio track through five image-plus-audio models, in sync and unranked. The session's checkpoint grid shows the difference that matters for this page: only the models that take a text direction act out the whistle.

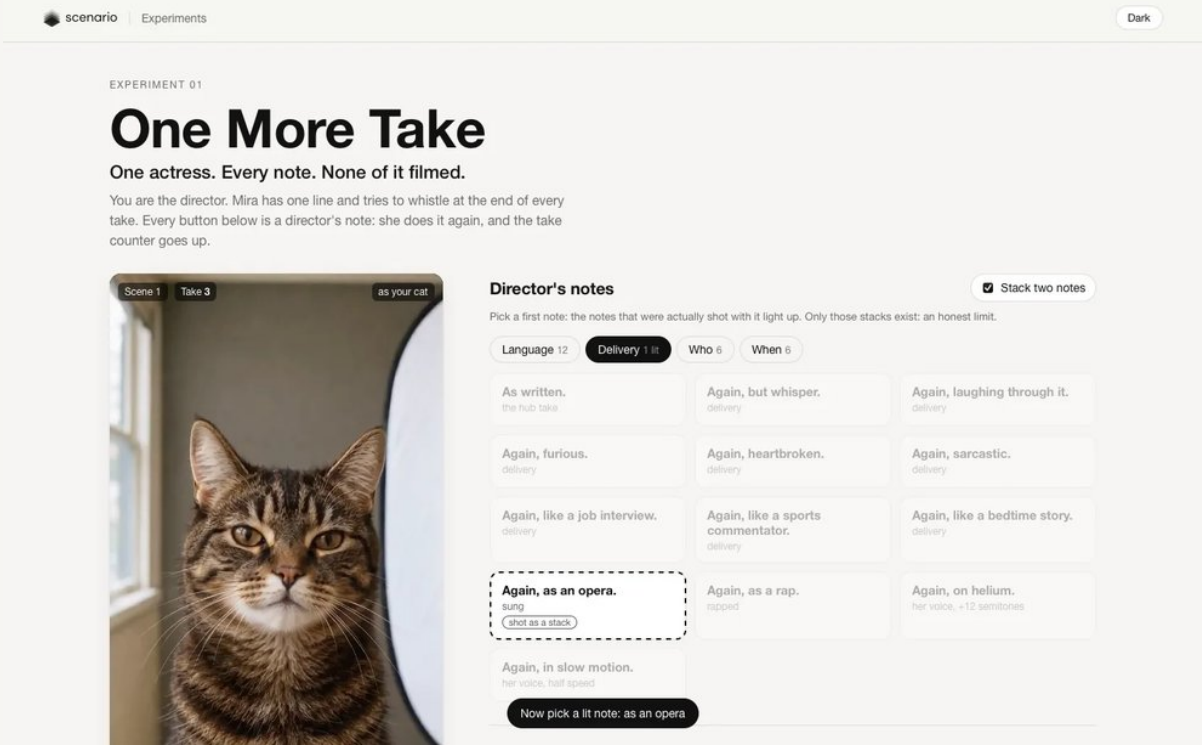


Mouth checkpoints at fixed times, from speech to the two air bursts of the whistle.

P-Video Avatar	36 CU · 47 s
Most expressive; acts the whistle; smiles through the pauses.	
Minimax H3 Max Lip Sync	226 CU · 28 s
Precise mouth shapes and lively brows; no whistle.	
Sync-3 Avatar	395 CU · 366 s
Calmost and most literal; closes on the pauses.	
Kling AI Avatar 2 Pro	255 CU · 331 s
Acts the whistle from its text direction.	
Omni Human 1.5	340 CU · 350 s
Most human body language; no whistle.	

From files to a set you can direct

After the full shoot the session wrote a first working page around its files: one data file with every take, its models, CU, model time and transcript; a monitor with a clapper and a take counter; four decks of notes; a way to combine two notes that lit only the pairs actually shot; a final cut of your last six takes kept in the address; and a recipe card and a replay of each take's measured generation. Nothing is generated live. It was checked in WebKit at desktop and phone sizes, light and dark. The published page reorganises that console around the same files.



The session's own screenshot of that first page: the cat on the monitor, the opera lit as the only pair shot with her.

CU BY MODEL, ALL THREE RUNS					
10,698	1,200	1,138	998	790	2,148
Seedance 2.5	Gemini Omni 1.1 Flash Edit (not on the page)	P-Video Avatar	Minimax H3 Max Lip Sync	Sync-3 Avatar	All other models
16,972 CU over 243 finished calls, 98 of them transcript checks.					