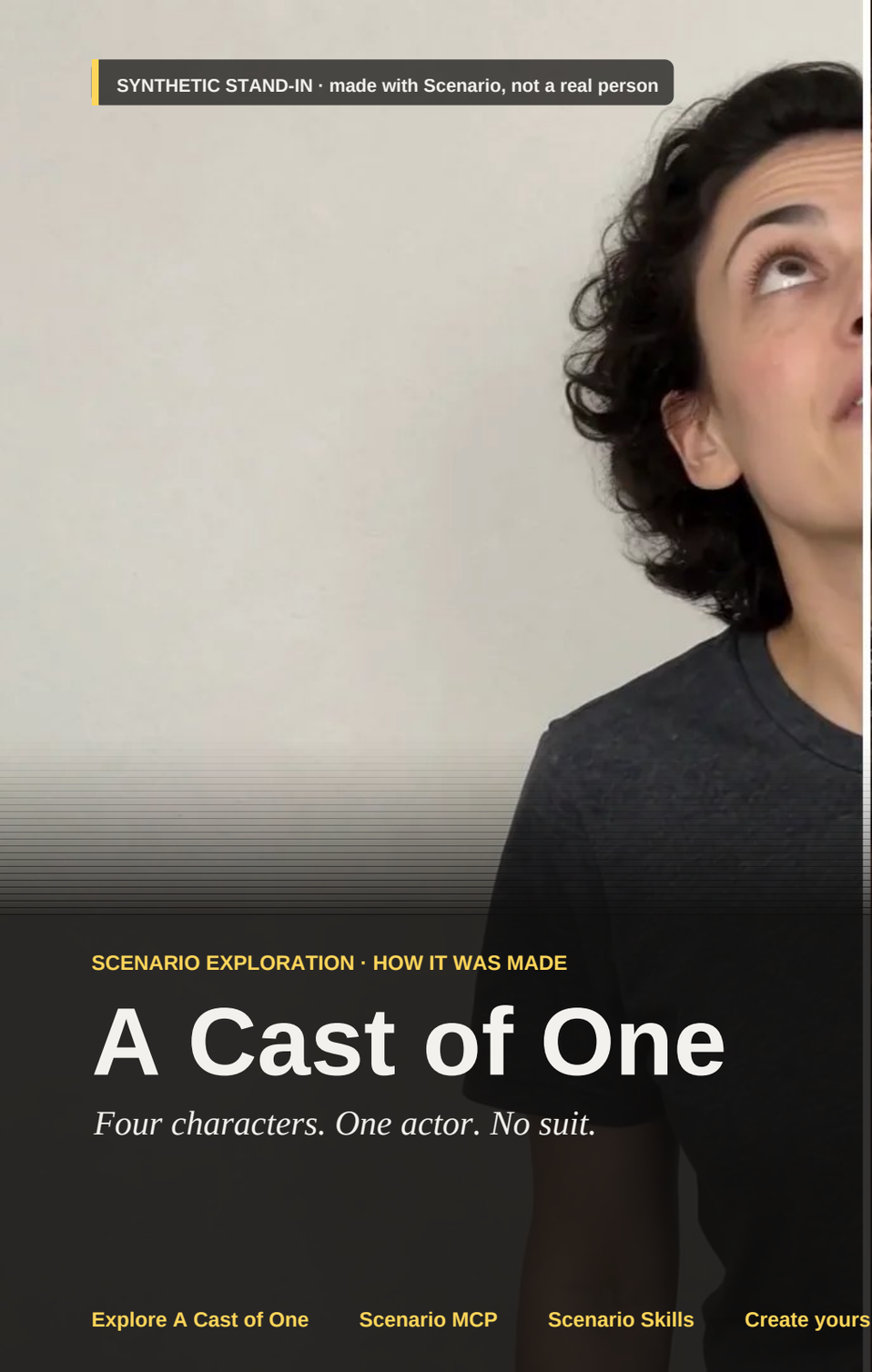
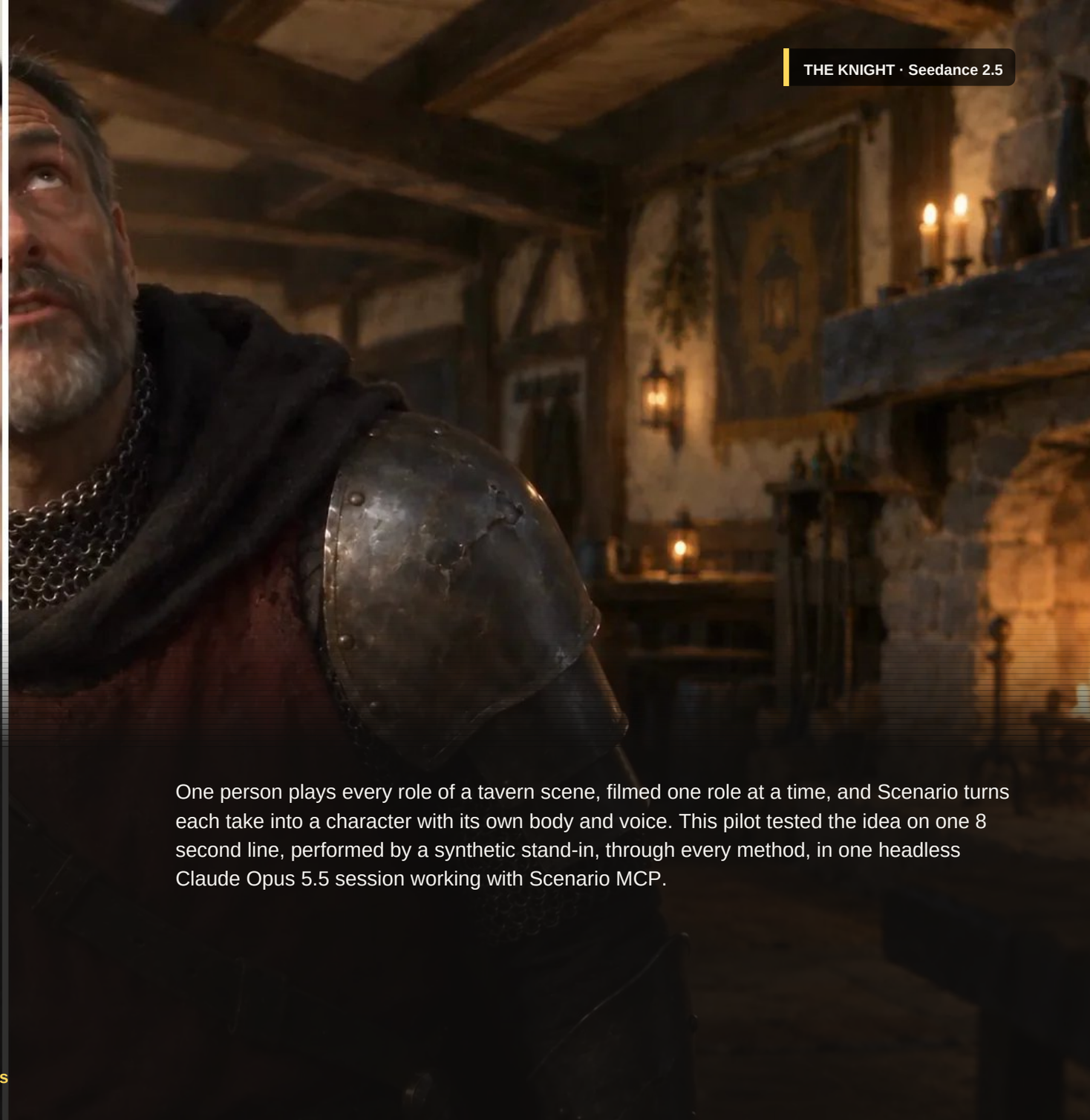




SCENARIO EXPLORATION · HOW IT WAS MADE

A Cast of One

Four characters. One actor. No suit.

One person plays every role of a tavern scene, filmed one role at a time, and Scenario turns each take into a character with its own body and voice. This pilot tested the idea on one 8 second line, performed by a synthetic stand-in, through every method, in one headless Claude Opus 5.5 session working with Scenario MCP.

Explore A Cast of One **Scenario MCP** **Scenario Skills** **Create yours**

THE BRIEF

One line, every method

The concept's own test, before anyone writes the final scene: one 8 second shot pushed through every method on the same performance, and measured for what holds. The real performer will be a team member filmed on a phone with written consent. That take does not exist yet, so the pilot ran on a synthetic stand-in made with Scenario, labelled as such on screen.

1 **THE LINE**

The knight's warning, written for the pilot. It closes the lips on the b of "blinked" and the m and b of "remembers", rounds them on "Quiet" and "word", opens them on "raven" and "down", and carries one big head move: the look up at the raven at 2.5 s.

2 **THE RULES**

Invented characters, names and places; no real person, no real voice, no voice cloning. Each model's live schema was read before its first use and every call was priced with a dry run before it ran. Costs are recorded in credits (CU) and seconds.

3 **THE CHECKS**

Every spoken take transcribed with Gemini 3.5 Transcribe, mouths judged on crops at eight word times, faces compared with the sheet, loudness and pitch measured on every voice, every clip watched as frame strips.

4 **THE RUN**

One prompt, one autonomous session. Every Scenario call went through Scenario MCP, most of them from the Scenario Blender Plugin's command-line client (Blender itself was not used): model search, schema reads, quotes, 35 generation jobs, uploads and a Scenario collection.



"Quiet. All of you. That raven on the rafter hasn't blinked since we sat down. It's listening... and it remembers every word."

The stand-in take at 2.7 s, the look up on "raven" (Minimax H3, 8 s, native audio). The pilot ran on 24 September, the page and a shoot kit followed on 25 September, and a Seedance 2.5 test on 27 September.

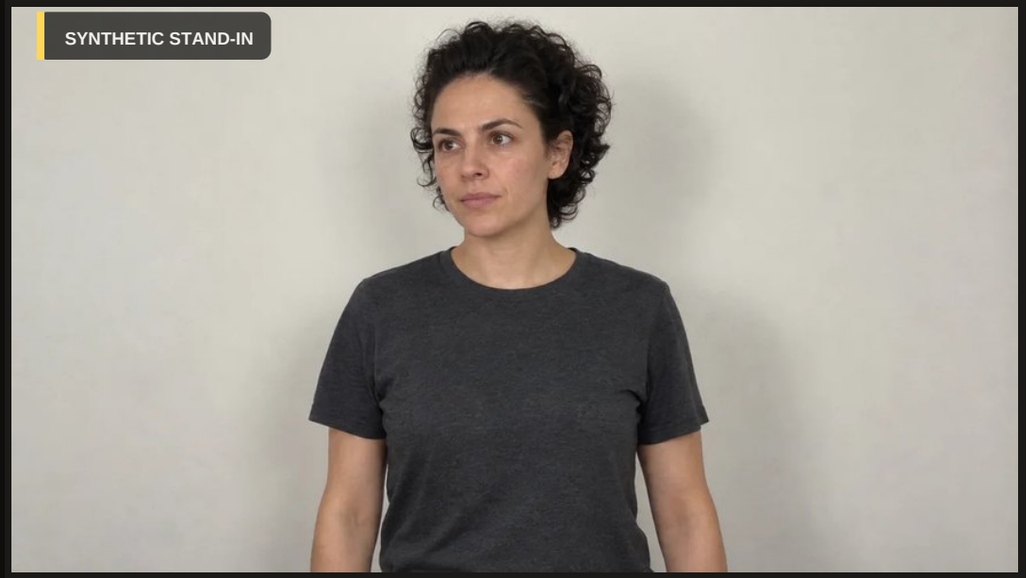
A stand-in, and four characters in one style

The stand-in

GPT Image 2.5 Sunburst painted the first frame as a smartphone self-tape: a locked tripod at chest height, a plain off-white wall a metre behind, two softboxes, natural skin with sensor noise and mild compression, an invented performer with her hands low and her eyes on an eyeline mark. Minimax H3 animated it into the 8 s take at 2K with native audio: a lean toward the eyeline, the look up on "raven", a frown on "remembers". The transcript came back with all 22 words.

The cast

One style block opens every prompt (a digital oil painting with realistic anatomy, visible brushwork and a warm candlelit palette), so the four sheets read as one production: a human-like knight, a cartoon goblin, a stylised wizard and a raven. Each faces the viewer, mouth closed, face evenly lit; beards stay short so the lips remain visible. The tavern is an empty plate with the middle ground left clear.



The stand-in's first frame (GPT Image 2.5 Sunburst): synthetic, not a real person.



The knight | *The goblin* | *The wizard* | *The raven* | *The tavern*

The Crooked Lantern, an empty plate. Every sheet and the set share one style block.

METHOD 1 · 24 SEPTEMBER

The knight's face: P-Video Replace

P-Video Replace swaps the performer for a character taken from identity references and keeps everything else: the timing, the body, the framing, the wall, the light and the lip motion. Because the voice changer keeps the take's timing too, the knight's mouth matched his new voice without a lipsync pass.



Face crops at five moments against the sheet: one reference (top row) and three references (bottom row).

One reference: fails

The knight came out younger, with fuller hair and no scar: the same man in every frame, but not the one on the sheet. The big look up at the raven was softened into a small tilt.

Three references: holds with a workaround

Two face close-ups edited from the sheet, front and three-quarter, and a prompt naming his age, wrinkles and scar brought back the rough short hair, the weathered skin and the scar. The look up stayed softened and the knight stayed against the plain wall.

METHODS 2 AND 3 · 24 SEPTEMBER

The goblin: mascot swap or motion transfer

The cartoon goblin shares nothing of the actor's proportions or mouth. The brief named Cartwheel Mascot for him and Kling V3 Pro Motion Control for the stylised wizard; when the mascot swap failed, the goblin went through Kling as well.



Cartwheel Mascot: on model, in a full-body wide.



Kling's input: the goblin placed in the tavern at the take's framing, his lips redrawn over his teeth.



Kling V3 Pro Motion Control: the take's performance kept.

Cartwheel Mascot: fails

It lit the goblin into the tavern convincingly, but the medium shot became a full-body wide with him small and to the left, the mouth stayed shut for the whole line and the look up was lost. It charged by the second of input, whatever the resolution.

Kling: holds

A Sunburst edit placed the goblin in the tavern at the take's framing, from the sheet, a frame of the take and the set. From that image and the take, Kling followed the look up, the head turns and the lips, at under an eighth of the mascot route's credits.

The wizard: holds

The same route, with the wizard placed in the tavern first, held on the stylised face too: the look up on cue, the frowns, the lips and the painted look. His hat brim shaded the upper face in some frames.

THE VOICE LANE

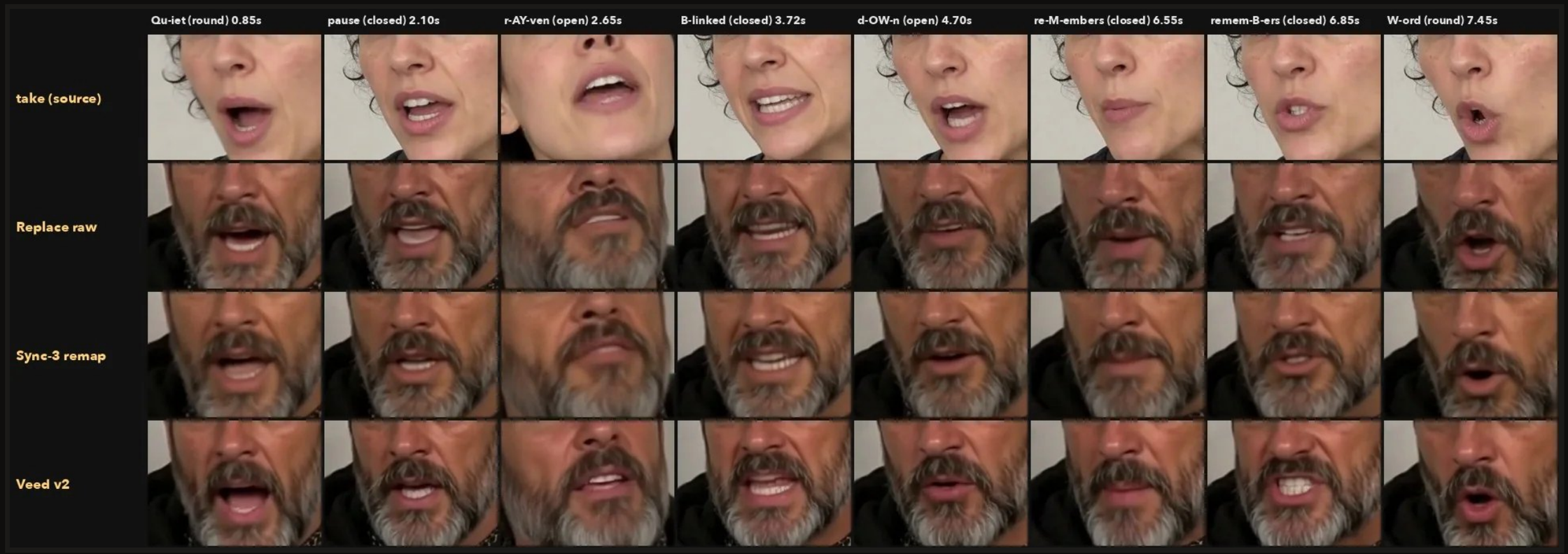
A new voice that keeps the timing

One preset voice per character

ElevenLabs Voice Isolator cleaned the take's audio, then ElevenLabs Voice Changer turned it into one preset voice per character. Words, timing and emotion survived: every output sits within 10 ms of the take and every transcript is word for word. The first three male presets all landed in one register, a median pitch of 124 to 134 Hz against about 210 Hz for the stand-in, so the goblin moved to a higher preset near 180 Hz. The knight and the wizard still share a register.

A lipsync pass only where needed

Where the body route keeps the actor's lips, the new voice already matches, so a lipsync pass is optional. On the human face, Sync-3 Lipsync in remap mode kept the face and closed the lips more cleanly on the b of "remembers"; Veed Lipsync v2 opened the mouth with teeth on that b and returned 25 fps. On the Cartwheel goblin, Sync-3 barely opened the mouth, and Sync React-1 opened it on the right words without restoring the lost performance.



Mouth crops at eight word times. Rows: the take, the one-reference knight, Sync-3 remap on it, Veed Lipsync v2 on it.

One prompt pattern, three roles, the same take

Three days later each role was re-made with Seedance 2.5 in reference mode from the same take: @video1 for the performance, camera, framing and timing only; the sheet as @image1, the tavern as @image2 and, for the knight, his two face close-ups. The prompt listed what must change (the performer becomes the character, the t-shirt the costume, the wall the tavern), locked the framing to the take's waist-up shot, named the mouth shapes at their words and asked for no audio, since the voices come from the voice changer. The three takes ran together, 256 to 267 seconds each.



The knight HOLDS

Head track against the take r 0.97 and 0.96 with no lag; the look up that P-Video Replace softened is back; the mouth right at 8 of 8 word times; the knight is in the tavern. It became his route.



The wizard HOLDS

r 0.98 and 0.98 with no lag, the mouth right at 8 of 8, the look up on cue. His three-quarter turns and a grimace came out softer than on Kling. It became his route.



The goblin DOES NOT HOLD

On model, but the head faced the lens where the actor turned away, the look up grew to a third of the frame and the track lagged by 0.3 to 0.5 s (r 0.52 and 0.49). The mouth was wrong at 2 of 8 word times. Kling stays his route.

THE REVEAL

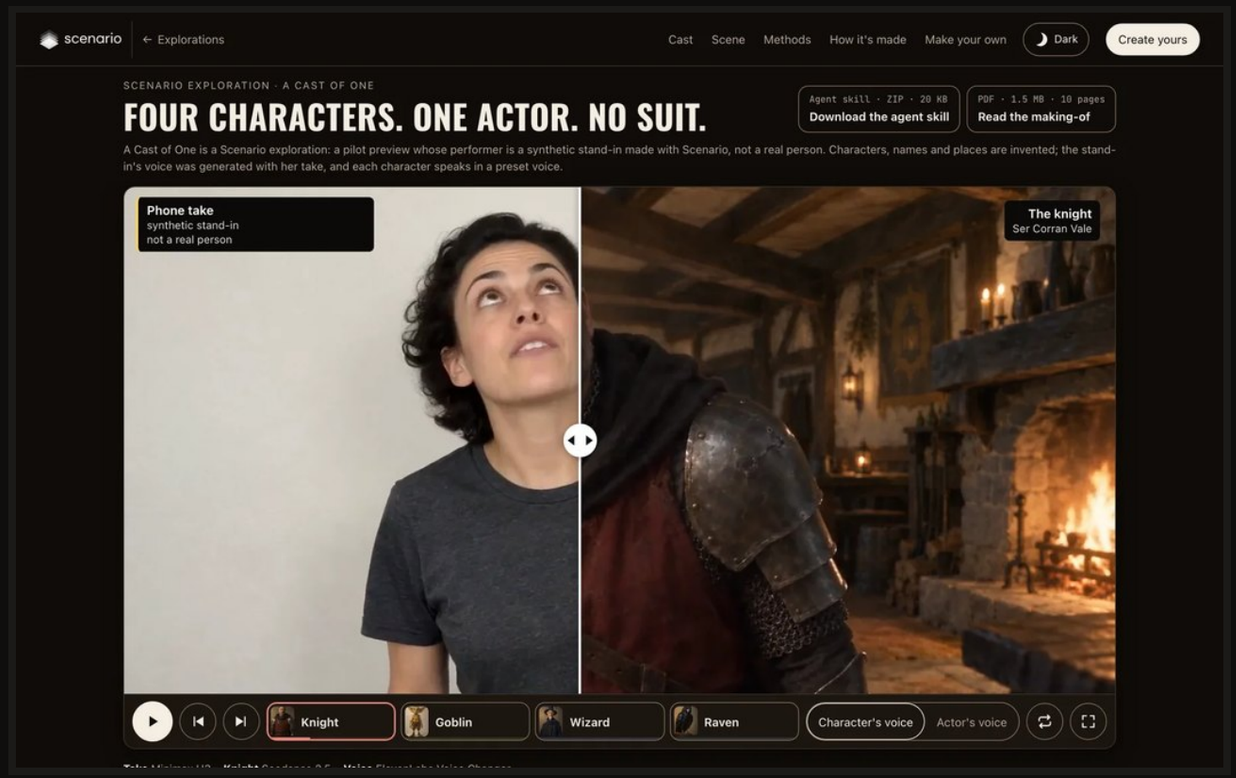
The same frames, from take to character

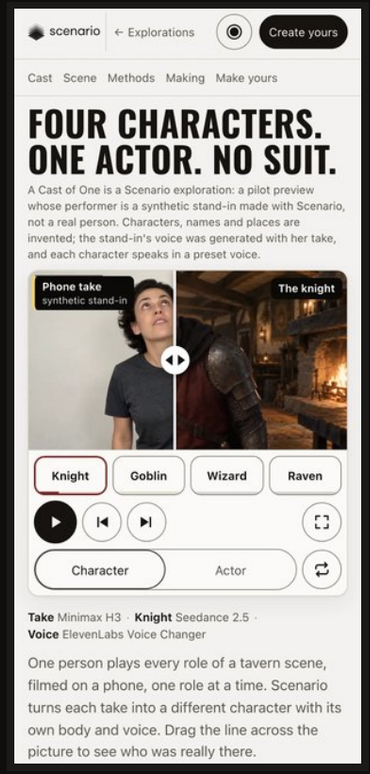
The same frames

The reveal slider needs the character and the take to line up frame for frame, and every route that held kept the take's timing. The edit conformed every output to 1920x1080, 24 fps and 8.0 s: P-Video Replace came back at 1920x1024 (an overlay test showed a resize, not a crop), Kling at 1920x1088 and 30 fps, Cartwheel at 1280x736 and three frames short. An ffmpeg wipe between the take and the knight then lined up on the same frames.

The page

The session built the page itself in plain HTML, CSS and JavaScript, with no external service and no model call: the scene with the pilot line playable in any role; the cast with the reveal slider, a voice switch between the actor and the character, and live captions from the transcript's word timestamps; "the line around the table"; the comparisons that decided the routes; and a replay of the measured pipeline at 20 times speed. A WebKit journey at 1440x900 and 390x844, in light and dark, checked it.





The cast player on scenario.com at the look up: the stand-in on the left of the divider, the Seedance 2.5 knight on the right.

WHAT HELD

Verdicts, the raven and the real take

Stand-in reads as a phone take	HOLDS
Knight, P-Video Replace	HOLDS WITH A WORKAROUND
Knight, Seedance 2.5	HOLDS
Goblin, Cartwheel Mascot	FAILS
Goblin, Kling V3 Pro Motion Control	HOLDS
Goblin, Seedance 2.5	DOES NOT HOLD
Wizard, Kling V3 Pro Motion Control	HOLDS
Wizard, Seedance 2.5	HOLDS
A different voice per role	HOLDS WITH A WORKAROUND
Lipsync pass	HOLDS WITH A WORKAROUND
Reveal slider	HOLDS
The raven, Minimax H3 Max Lip Sync	HOLDS

The real take

Everything here was measured on synthetic footage. The real version needs a team member filmed on a phone, written consent covering the raw take and their own voice on the page, preset voices only, and the whole grid re-run on that take. The shoot kit sets the rest: a locked phone at chest height, 24 fps, the same waist-up frame for every role, a plain wall, two soft lights, hands away from the mouth, turns to three-quarter, 10 to 15 s takes, a clip-on mic and eyeline marks.



Old Sable speaking the line: Minimax H3 Max Lip Sync from its closed-beak still.

The raven

The raven was first left as a feasibility note and one still. A revision the same day made it speak: its still redrawn with the beak closed (GPT Image 2.5 Sunburst), its own preset voice made from the actor's take (ElevenLabs Voice Changer), and the beak moved to that voice by Minimax H3 Max Lip Sync, the best of six takes on the same voice. The beak opens on the syllables and shuts in the pauses and on the m and b; the head stays in profile. The same revision redrew the goblin's lips over his teeth and re-made both his takes.

THE NUMBERS

35 jobs, 6,949 CU

Every job as recorded in the session's ledger, with the credits (CU) it was charged. About 2,490 CU of the pilot went to routes that failed or ran only for comparison; the Seedance test added 3,270 CU. A revision after review then added 34 jobs and 3,776 CU (18 generations and 16 scoring checks): 69 jobs and 10,725 CU in all.

MODEL	USED FOR	JOBS	CU
GPT Image 2.5 Sunburst	Sheets, set, first frame, framed inputs, close-ups, raven	11	135
Minimax H3	The stand-in take, 8 s with native audio	1	208
ElevenLabs Voice Isolator	The take's voice	1	4
ElevenLabs Voice Changer	Four voices, one of them set aside	4	16
Gemini 3.5 Transcribe	Words check on the take and every voice	5	10
P-Video Replace	The knight, one and three references	2	192
Cartwheel Mascot	The goblin, the rejected route	1	1,728
Kling V3 Pro Motion Control	The wizard and the goblin	2	404
Sync-3 Lipsync	Remap on the knight twice and the goblin once	3	669
Veed Lipsync v2	The knight, compared	1	112
Sync React-1	The goblin's mouth	1	201
Seedance 2.5	The three roles, reference mode	3	3,270
Total, three runs of one session		**35**	**6,949**

Time

The session took about 60 minutes for the pilot (32 jobs in about 50 minutes of wall time, several in flight at once), 33 minutes for the page and the shoot kit with no generation, and 22 minutes for the Seedance test.

The real scene

For the real 45 s scene on these routes, the pilot estimates about 17,200 CU with retakes, a half-day shoot and about one working day after it.

Explore the page, download the agent skill, and film your own cast of one.