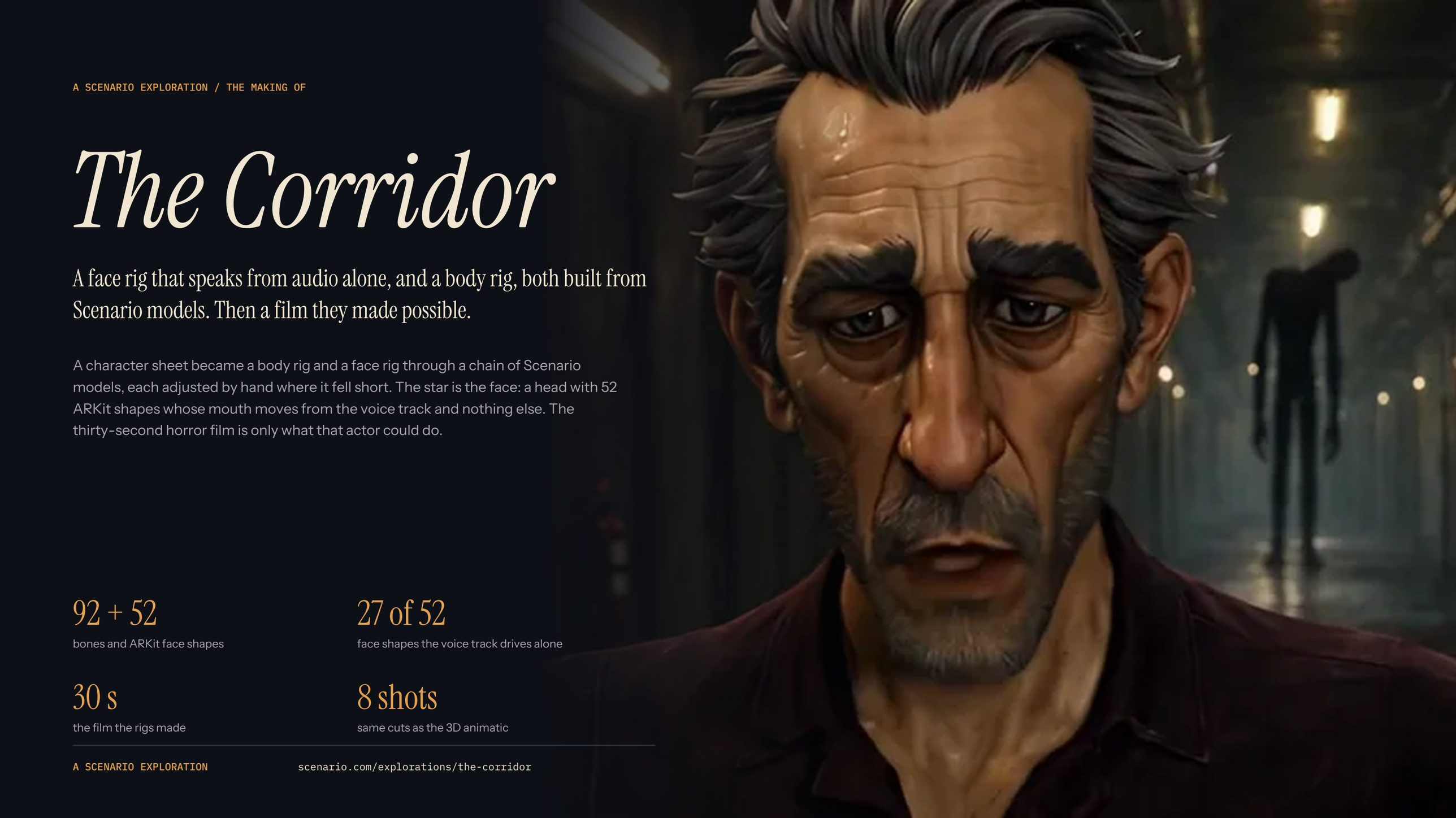
A SCENARIO EXPLORATION / THE MAKING OF

The Corridor

A face rig that speaks from audio alone, and a body rig, both built from Scenario models. Then a film they made possible.

A character sheet became a body rig and a face rig through a chain of Scenario models, each adjusted by hand where it fell short. The star is the face: a head with 52 ARKit shapes whose mouth moves from the voice track and nothing else. The thirty-second horror film is only what that actor could do.



92 + 52
bones and ARKit face shapes

27 of 52
face shapes the voice track drives alone

30 s
the film the rigs made

8 shots
same cuts as the 3D animatic

The rigs are the work. The film is the result.

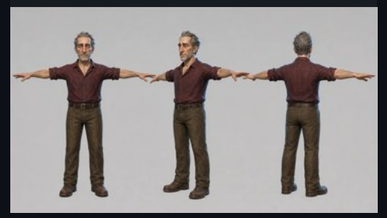
This exploration is about building an actor that can perform: a body rig and a face rig made from Scenario models and adjusted by hand until they hold up on camera. The proof is the face, which speaks from a voice track alone. Once the actor could act, a 3D animatic fixed every cut and move, and video models only finished the picture.

01

GPT IMAGE 2

Draw

Four character sheets. One man is chosen.



02

MESHY 7.1 + ASTRA 3D

Build

Body and head become 3D meshes, then one character.



03

ASTRA 3D RIG

Rig

92 bones for the body, 52 ARKit shapes for the face, then hand repairs.



04

UNREAL ENGINE 5.8

Act

Cameras, lamps and retargeted clips. A voice track moves the face by itself.



05

SEEDANCE 2.5 + MINIMAX H3

Finish

The animatic is the reference. The models add the surface.



06

SYNC LIPSYNC

Repair

Shot-by-shot fixes and a lip pass on approved shots.



Which model did what.

Every step of the actor went through a Scenario model. This is what each one did, and what had to be adjusted by hand.

STEP	MODEL	WHAT IT DID	WHAT WE ADJUSTED
Character sheets	**GPT Image 2**	Four sheets of the man; a body with a blank sphere head, and the head alone	The head was drawn with the mouth slightly open, so the jaw would skin cleanly
Body and head meshes	**Meshy 7.1, image to 3D**	Body and head built apart, 4K geometry, 4K PBR textures	Fitted in Blender by the director; nothing was cut at the neck
Decimation and rebuild	**Claude Opus 5.5 3D, then GPT-6 Astra 3D**	Opus smeared the head texture; Astra rebuilt both meshes keeping the original UVs and colors	Body 150,000 faces, head about 120,000; loose shirt islands checked before rigging
Body and face rig	**GPT-6 Astra 3D**	92 bones with Unreal Mannequin names, fingers, twist bones, IK, and exactly 52 ARKit face shapes	Jaw weights repaired in space; eyes, teeth and tongue kept as separate meshes
Body motion	**Uthana, Meshy and Cartwheel text to motion**	Three services compared; none was used as is	Epic walk and idle retargeted by bone name, the panic pose built by hand
Voice	**Gemini 3.8 Flash TTS**	The spoken lines	A script turned loudness and vowel color into jaw and mouth shapes








A man, on paper first.

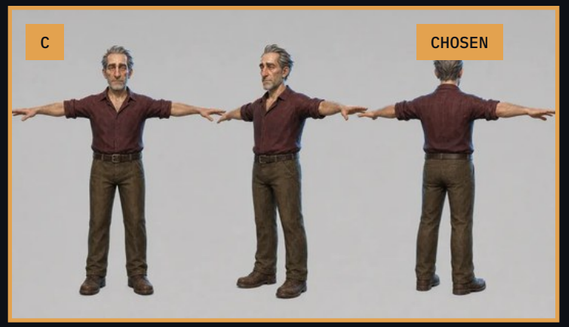
GPT Image 2 drew four T-pose character sheets, each with a front, three-quarter and back view, so the cut lines between head, torso and legs stayed readable. The man with the crooked nose and the burgundy shirt was chosen: wrinkles, heavy lids and bushy brows are raw material for a face rig.



Mustard shirt, forties



Teal shirt, thirties



Burgundy shirt, fifties



Orange shirt, twenties



The body, pushed toward a painted look.

The chosen sheet was redrawn toward a painted look, mixing the weight of Arcane with the graphic clarity of Spider-Verse, with human proportions and simple hair.



The head, drawn on its own.

The head was a separate image with the lips slightly parted, to give the jaw room before the rig closes it to a neutral mouth with separated lips. The body went to 3D with a plain sphere for a head.

Two meshes, one man.

Building the head and the body separately kept each at a useful resolution and let the face be rigged on its own.

Image to 3D

Meshy 7.1 Image to 3D built the body from the sphere-headed image and the head from the head sheet, with 4K geometry, 4K textures and PBR maps.

Lighter, not smeared

A first polygon reduction with Opus 5.5 3D smeared the head texture. Astra 3D jobs that keep the original UVs and bake only a normal map were approved: a body of 150,000 faces and a head of about 120,000.

One character

A cut-and-join assembly left a visible seam at the collar. The final assembly overlaps head and body untouched and matches their skin tones.



Raw body from Image to 3D, with the sphere head



The assembled character in Unreal



The head, rigged and lit

A skeleton for the body, 52 shapes for the face.

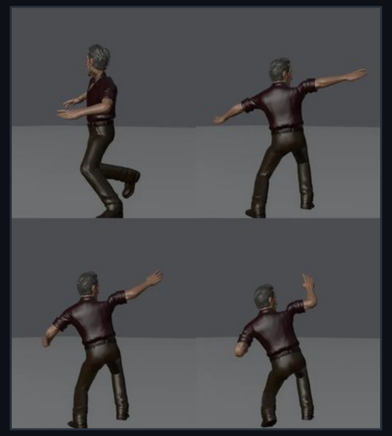
An Astra 3D rig job produced 92 bones with the Unreal Mannequin naming: 64 core and finger bones, 16 twist bones, 7 IK bones, the jaw and both eyes, and two thumb metacarpals. The head carries exactly the 52 ARKit blendshapes, with separate eyeballs, teeth and tongue.

Body 147,778 triangles, no unweighted vertices, at most four influences per vertex, weights normalized.

Face 52 ARKit keys on a head of 114,887 triangles, jawOpen at 62 mm, blink at 8 mm. Eyes, teeth and tongue are separate meshes.

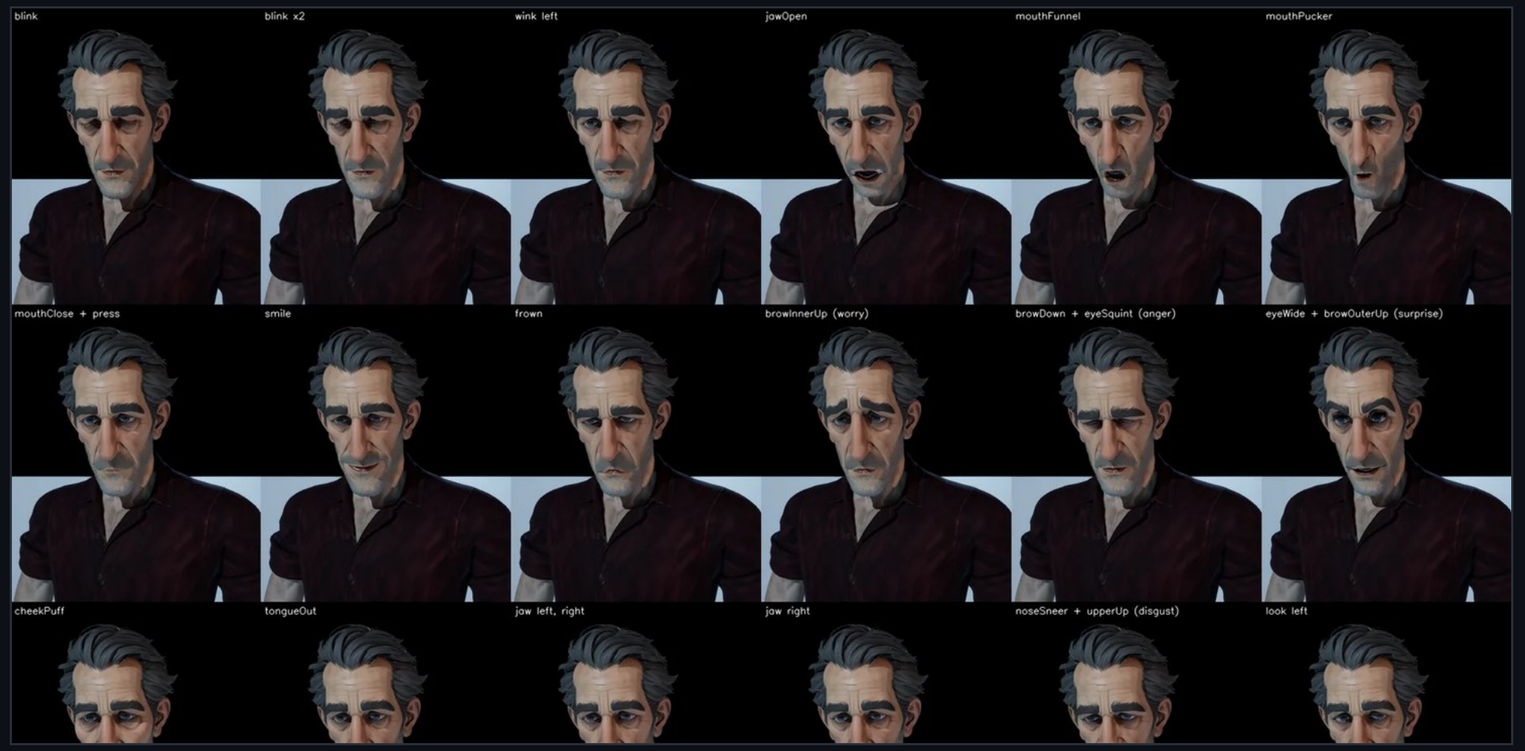
Names Mannequin bone names, so Unreal animation can be retargeted by name.

Test clip



Face tests





Every ARKit shape driven on its own in Unreal, from mouth close to tongue out.

Making the face behave.

Importing into Unreal 5.8 gave a skeleton with all 52 morph targets but no textures, so the PBR materials were rebuilt from textures exported from the rig file. Then the face needed three fixes.

01

Jaw weights

The jaw skin weights were smoothed by spatial distance rather than mesh connectivity, so beard and hair shells move with the chin while the lip seam stays sharp.

02

Jaw amount

Speech opens the jaw by a fraction of its range, lowered across versions from 0.62 to 0.36, and always vertically, never stretched sideways.

03

Torn eyes and lids

Duplicate keys at the same time in the ARKit curves made the eyelids overshoot and the face tear. The curves are now sampled every frame, summed and clamped, and a blink cancels squint and wide.



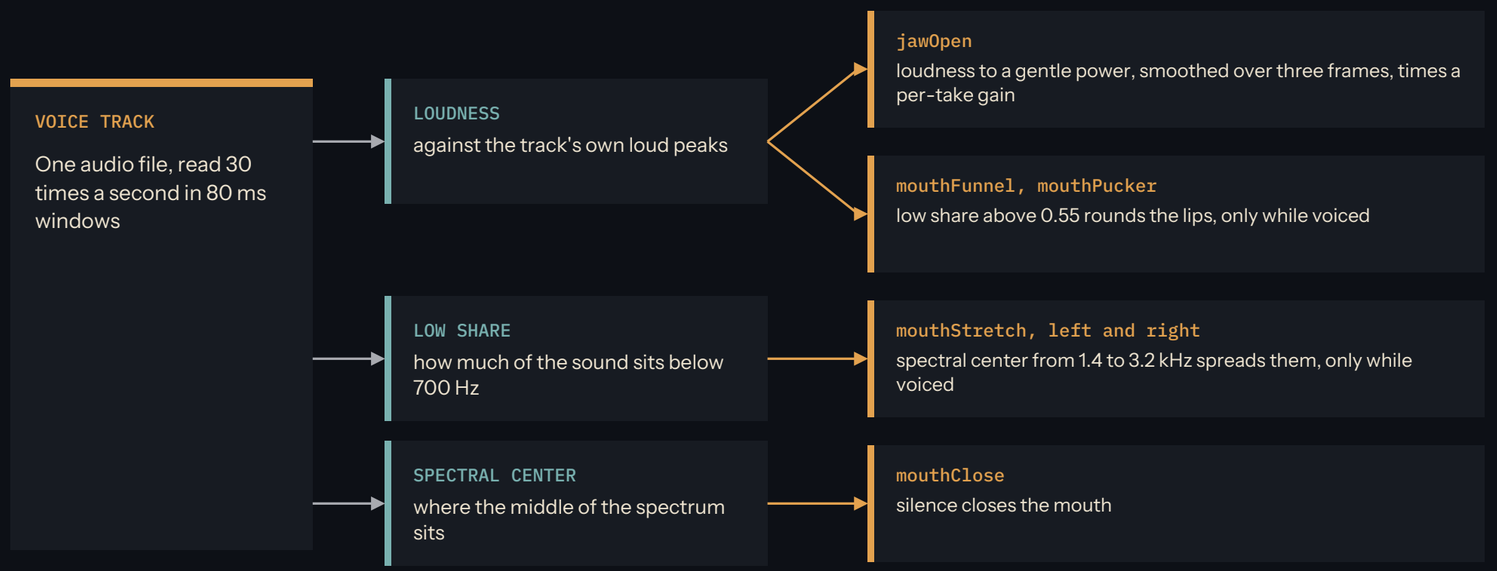
Jaw skinning before (left) and after (right)



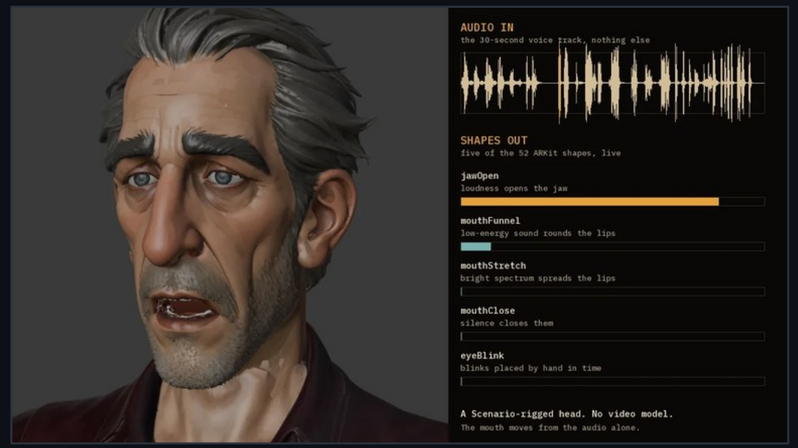
Jaw amount, first pass and second pass

Audio in. A face out.

Eight English lines were voiced with Gemini 3.8 Flash TTS in a hoarse, older voice whose pitch climbs from calm to panic; six remain in the thirty-second cut. A script reads the voice track and writes ARKit curves for the head's shape keys. Nothing else moves the mouth: no video model, no hand animation of the lips.



From the same numbers: the lower lip drops and the upper lip lifts with the jaw, the cheeks squint a little, the brows rise slowly on whole phrases and are pulled down while the jaw is open, and the gaze drifts. Placed by hand in time: twelve blinks.



A still of the demo video on the exploration page: the real rig on the left, the voice track and five live shape values on the right.

The mouth follows the sound.

Seven frames of the real rig at different moments of the same voice track, with the live value of four shapes under each and the moment marked on the waveform. They come straight from the demo video on the exploration page.



9.7 s

jawOpen 0.42: a loud, open vowel.

10.2 s

mouthFunnel 0.68: low, round sound, lips pushed forward.

19.9 s

mouthStretch 0.50: bright sound, lips pulled wide.

6.1 s

Silence: jawOpen 0.01, mouthClose 0.35, lips together.

What it measures, and what it cannot do.

The numbers below come from the curves behind the demo video (jaw gain 0.50). A rule set on top of a good rig is not speech recognition, and these are its edges.

27 of 52
shape keys driven by the voice track

0.11
jaw average, as a share of its range

0.43
jaw peak at a gain of 0.50

0.36 to 0.62
jaw gains used across the film's takes

1 frame
the mouth is placed ahead of the sound

12
blinks, placed by hand in time



The same rig at a loud vowel and in a held silence.

It hears color, not words

There are no phonemes. The mouth reads loudness and the shape of the spectrum, never the words.

Closed sounds are not closed

A p, a b or an m does not close the lips on purpose; the mouth closes only in silence.

Judged by eye

The result was reviewed by watching it against the voice. No lip-error figure is claimed.

A gain per take

The jaw gain was lowered from 0.62 to 0.36 over the early versions and set per take.

Not the finished mouths

The mouths in the finished film come from the video model, plus a lip-sync pass on approved shots. The next pages show why.

The axes were not what they looked like.

Much of the time went into learning how this rig bends, then teaching the engine to respect it.

The arm axis is a twist.

A first guard pose put the hands behind the shoulders because the upper arm's X axis is the bone's twist. In this rig the upper arm lifts forward on Y and the elbow folds on Z, mirrored on the right side.

Neck and spine.

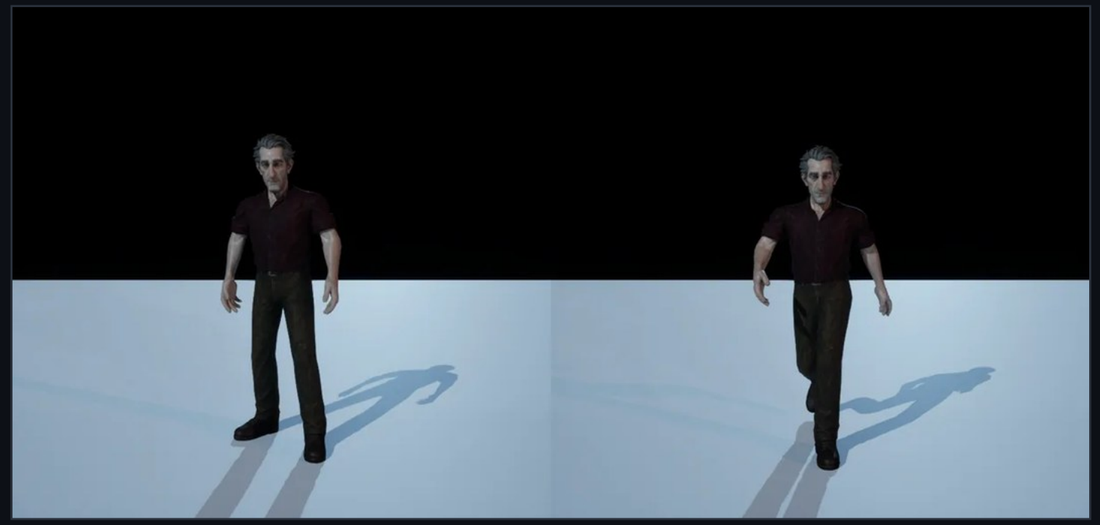
The neck turns on local Y and nods on X; Z tilted the head down. The spine leans forward on negative X.

Retarget by bone name.

Epic's idle and walk retargeted automatically. The text-to-motion clip from Uthana leaned and fell into a T-pose in the automatic retarget, so a script retargets by bone name instead and drops the hip translation curves.

What the engine dropped.

The FBX import brought no textures, so the PBR materials were rebuilt. 4K textures filled the video memory and made patchy renders, so everything went to 2K except head skin and eyes.



Epic idle and walk retargeted onto the rig.

Borrowed clips, handmade acting.

No performer was motion-captured for the final film. The body comes from retargeted clips, shaped and extended by hand.

Retargeted clips

Epic's idle and walk were retargeted onto the rig. The walk is sampled at 0.55x speed with root motion, and head glances are added on top.

A three-way bake-off

Three text-to-motion services were compared on the same rig: one gave the best gestures, one used a simpler skeleton, one gave introspective head-down acting. None became the film's walk, but the tests shaped the plan.

The last pose, by hand

The guard pose is built on the rig by hand: crouched, leaning forward, both hands open. The axis lesson on the previous page is why it works.



The same prompt through three text-to-motion services



The final guard pose in the animatic

A new skeleton, onto the same rig.

The film uses library walks and a hand-built pose, not motion capture. Afterwards, Viggle's text to motion and video to motion models were tested on the same rig. They export a Mixamo skeleton.



Five test clips on a gray mannequin: nervous walk, head tilt, flinch, guard stance, and capture from the film's dolly shot.

One profile, all bones

A Mixamo profile was added to the retarget script. A test clip landed on all 52 mapped bones, none missing, and the bone directions matched the source exactly. GLB opens in Blender; the FBX export is ASCII, which Blender cannot import.

The roll

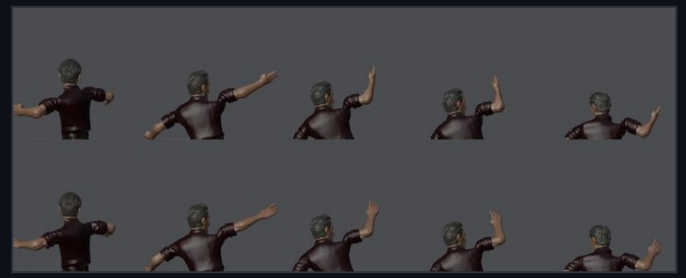
Directions matched but both hands were rolled a quarter turn, 83 degrees by the palm axis, because the rig's rest pose differs from the T-pose by about 84 degrees. Adding that twist to the whole arm chain brought the palm error to 1.4 degrees.

What was tested

Four text prompts (a nervous walk, a head tilt toward a noise, a flinch with hands to the mouth, a guard stance into a sprint) and one video clip taken from the film's own dolly shot.

What it did not do

None of this is in the film. The first second of the video clip is a close-up, and video capture needs the whole body in frame.



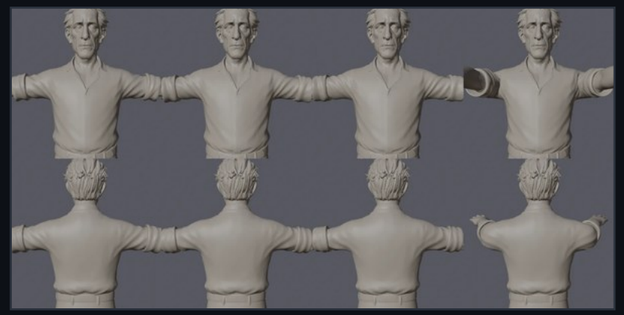
Hands before (top) and after (bottom) the roll fix, same frames.

The shoulder dents when the arm goes high.

With the arm raised, the skin between neck and shoulder dents. The retarget is not the cause: the collarbone directions matched the source exactly. The skin weights are.



The same frame after each attempted repair: creases, flat patches, and an arm that left the shoulder.



Static skin tests (neutral, arms forward, arms to the side) show no collapse in these poses.

The cause, in the weights

At 20 centimeters from the center line the collarbone carries 13 percent of the skin weight against 56 percent for the top chest bone; at 24 centimeters, 10 against 34. When the arm rises the collarbone moves and the skin tied to the spine does not.

Try 1

Moving up to 80 percent of the chest weight to the collarbone between 9 and 33 centimeters, then smoothing: creases and hard folds in the trapezius.

Try 2

55 percent over a wider band with more smoothing: the smoothing flattened the sculpted shirt folds and left patches, with the creases still there.

Try 3

Halving the collarbone rotation in the retarget: the arm, placed in world space, left the shoulder and passed through the torso.

Status

Not repaired. The film's shots keep the arms low. Options are painting the weights by hand or asking a rig model for new shoulder weights.

The animatic is the director.

Unreal rendered the performance as a thirty-second animatic. The cameras stay inside the corridor, close on the face, and the whole corridor is never shown at the same time as his face. Lamps die one by one and a procedural dirt material rises over the walls as the film goes on.



Frames from the thirty-second animatic render. Flicker, grain and letterbox were added in post; the finishing pass later removed the flicker on purpose.

0.0 to 2.8 s	Extreme close-up of his eyes
2.8 to 6.0 s	His forehead, from above
6.0 to 8.0 s	From behind, walking
8.0 to 13.5 s	Top-down, tracking him
13.5 to 16.5 s	Over the shoulder, he glances back
16.5 to 19.5 s	Frontal close-up, wide with fear
19.5 to 25.5 s	Dolly back, lamps dimming
25.5 to 29.9 s	Guard pose, the last lamp dies

The animatic above. The picture below.

The same instants of the film: framing, pose and timing come from the animatic; surface, light and acting come from the finishing pass.



Eyes	Brow	Behind	Top-down	Shoulder	Fear	Dolly	Guard
1.2 s	4.0 s	7.0 s	9.7 s	14.8 s	17.3 s	22.0 s	27.0 s

Final frames come from the final 1080p cut. The shots from 6 to 13.5 seconds were finished with MiniMax H3 in the last round; the others come from the Seedance 2.5 finishing passes, upscaled to 1080p.

The color came from the animatic.

In the top-down shot the hair went brown, gray, blue-white and black and the skin turned pale. Writing the colors into the prompt did not stop it. The cause was the animatic itself: its lighting changed the hair and skin color from frame to frame, and the model copied it faithfully.



Three changes fixed it.

Finished keyframes with the right colors were added as references. The animatic was converted to grayscale, so the video carries motion and framing but no color. And for shots where the geometry had to stay locked, Seedance's edit mode ran on that grayscale video. The last row shows the shot redone in the last round (pages 20 to 22).

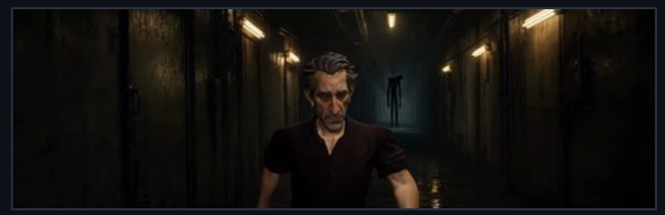
One corridor, eight shots.

Each shot was finished on its own, and the model reinvented the walls every time. A corridor bible fixed it: one wall finish, a steel door with plates and a red fire-hose cabinet on the left, a ventilation grille and a black pipe on the right, a regular row of warm amber tubes, wet floors. The same text and the same two reference stills went into every corridor shot.



Five corridor shots finished with the bible, one row each. The final cut keeps the best take of each shot, except 6 to 13.5 seconds, which was redone.





Edit mode locks the set

With the grayscale animatic as the video to edit, the camera and the architecture stay exactly where the animatic put them, so the background no longer drifts.

The lamps make sense

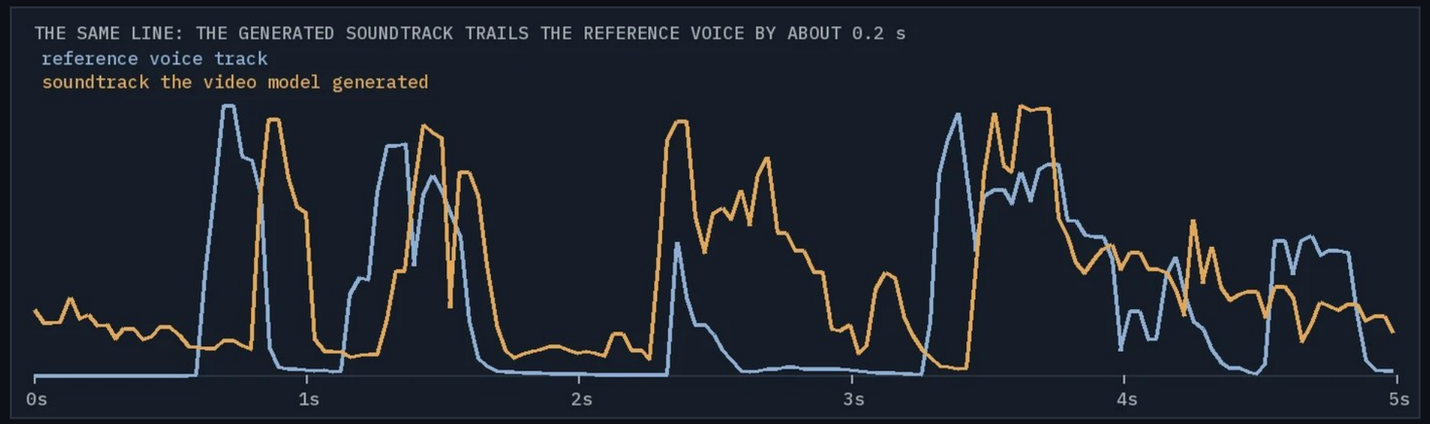
The dolly shot was relit with many more amber tubes, so the lamps that die later in the film have something to die from.

Where it fell short

On plain, featureless walls the model painted large cloud-like blotches that differed from the walls of the other shots. The two worst shots, from 6 to 13.5 seconds, were finished again in the last round (pages 20 to 22).

A dedicated lip-sync pass.

In several finished shots the mouth kept moving after a line had ended. The soundtrack the video model generated trailed the reference voice track by about 0.2 s, and measuring the mouths directly on stylized faces was too noisy to settle an exact figure.





The same instant of the top-down shot at about 11 s, before and after the pass.

A mouth-only rewrite

A lip-sync model rewrote the mouth area of the finished picture to match the voice, leaving everything else alone.

Used where it earned its place

The pass was kept for the shots at about 10 to 12 s and 16 to 19 s, which was where the face is clearly visible. It ran again on the reworked top-down shot, and only its 10 to 12 s window was kept. The over-the-shoulder profile at 14 s looked worse with it and was dropped.

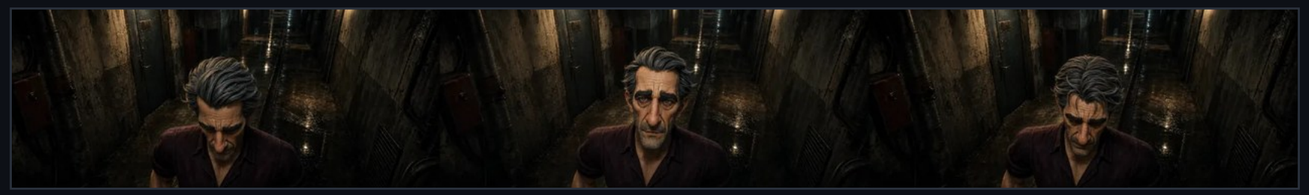
The frontal shot, redone

The frontal close-up was regenerated to follow the animatic's head snap, with fear added, and then lip-synced.

One corridor per shot.

The finished keyframes looked right, yet the first new tries at the shots from 6 to 13.5 seconds still faded between corridors. The mistake was in the references. The first frame and the last frame of a shot were two separate pictures whose corridors did not match, so the model had to turn one into the other.





The three frames used as image references for the top-down shot: bowed, looking into the lens, lowered. One corridor.

The fix

Every frame of a shot now comes from one picture of the corridor. The start and end poses were GPT Image 2 edits of the 10.7 s keyframe, with the corridor locked and only the man changed: head bowed for the start, lowered again for the end.

The recipe

The grayscale animatic goes in as the video reference, for motion, timing and framing only. The three frames of the shot go in as image references with one role each: the start, the moment he looks into the lens, the end. Colors and walls come from the images.

Why it works

Given two different places, a model invents the way between them. Given one place and three poses of the man, there is nothing to invent but his movement.

Same inputs, two models.

Both models received the same grayscale animatic, the same three frames and the same prompt. Both followed the order of the animatic, head bowed, up into the lens, down again, with the finished look and without the earlier washed-out color. They differed in what they did to the corridor.



The top-down shot, five instants across each clip.



The walk from behind, 6 to 8 s: the animatic cut and the first two seconds of the MiniMax clip.

Seedance 2.5

A 5.0 s clip at 1470 by 630. The door, the fuse box and the vent grille stay where the first frame put them.

MiniMax H3

A 6.6 s clip at 2944 by 1248. The camera drifts a little: the door moves and a cracked wall appears.

The choice

The director chose MiniMax H3 after seeing both, and it finished the two reworked shots.

The walk from behind

A new keyframe: the narrow warm frame redressed with the corridor of the 10.7 s keyframe, and one MiniMax run. The model needs at least 5 seconds, so the 2 seconds of the shot were cut from a 5 second clip.

Spliced into the 1080p cut.

The two reworked shots replaced the old ones in the 1080p cut: the walk from behind at 6.0 to 8.0 seconds and the top-down shot at 8.0 to 13.6 seconds. The audio is the original, and only the 10 to 12 s window carries the new lip-sync pass.



The same instants of the previous cut and the final cut.

THE PROMPT, BOTH MODELS

Finish the director's animatic @video1 as the final film shot. Follow @video1 exactly for the movement, timing, framing, camera angle and the man's head and body motion; it is a grey blocking video, so take only its motion and composition and ignore its grey colour, flat walls and lighting. The look, colour, materials and the corridor come only from the three finished stills, which show the same corridor from the same camera: @image1 is the opening pose with his head bowed, @image2 is the middle moment where he looks straight into the lens, @image3 is the closing pose with his head lowered again. Keep the corridor identical in every frame, with the steel door and red fuse box on the left, black cables and pipes, the slatted vent grille on the right, the wet reflective floor and the warm amber tube lights; nothing in it may change or fade. Rich cinematic image, deep saturated amber and teal, high contrast, crisp skin and hair detail, ash-grey hair, burgundy shirt, no washed-out or grey colour, no extra people.

WHAT IS STILL IMPERFECT

- The corridor still shifts a little from frame to frame in the MiniMax shots.
- The hair in the walk from behind is lighter than in the other shots.
- The pace follows the animatic roughly: the 6.6 s top-down clip was compressed to the 5.5 s of the shot.
- Mouths were judged by eye, and the lip-sync pass runs only in the 10 to 12 s and 16 to 19 s windows.

Block it yourself.

Rig an actor, rough out the performance in 3D, keep the animatic as the contract and let the model finish only the surface.

See the film and the side-by-side ->

Scenario MCP, to run it from your agent ->

Scenario Skills ->

MADE WITH

Character sheets GPT Image 2	Meshes and rig Meshy 7.1, Opus 5.5 3D, Astra 3D	Body motion Uthana, Meshy, Cartwheel
Voice Gemini 3.8 Flash TTS	Finish Seedance 2.5, MiniMax H3	Lip sync Sync Lipsync 2 Pro
Capture tests Viggle	Stage Unreal Engine 5.8, Blender	Upscale Topaz Video Upscale

