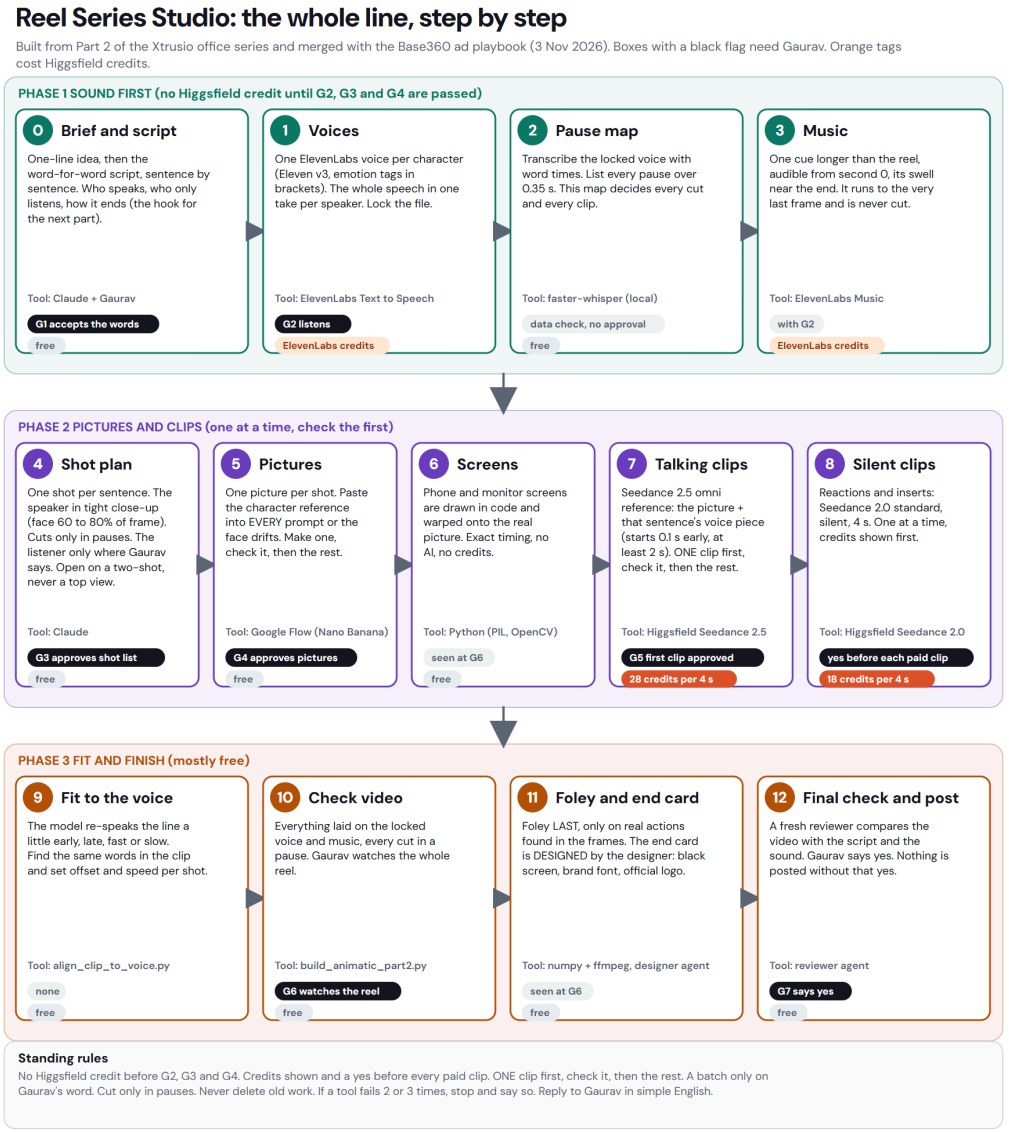
Reel Series Studio: architecture and lessons

How we make a short vertical drama reel: sound first, one clip at a time, cuts only in pauses. Built from Part 2 of the Xtrusio office series, the Base360 ad playbook and the ReelShort study (3 November 2026). Pictures first, then the written architecture, lessons, cost ledger and style rules.

The whole line



How ONE talking shot is made

Example from Part 2: Martin says "Seriously, who even made him CEO?" Same pattern that worked for the Base360 ad, plus the timing fit.

1 Find the sentence

In the pause map: first word starts 16.54 s, last word ends 18.38 s.

pause map

2 Cut the voice piece

From 0.1 s BEFORE the first word (16.4 s) to just after the last (18.6 s). Pad with silence to 3 s: the model needs 2 s or more.

ffmpeg

3 Picture

Tight close-up in Flow (face about 75% of the frame), made with the character reference pasted in the prompt.

Google Flow

4 One Seedance call

Seedance 2.5, mode omni_reference, 720p, 9:16, 4 s. Picture = image reference, voice piece = audio reference. 28 credits. Check it BEFORE making the others.

Higgsfield

5 Fit to the voice

Find the same words in the clip. Result for this shot: starts at clip second 0.00, speed 0.853 (the clip speaks slower than the voice). The check video takes these two numbers.

align_clip_to_voice.py

6 Cut in the pause

The shot starts at 16.54 s (the pause before it) and the next cut waits for the next pause.

check video

What went wrong in Part 2 (do not repeat) ~330 of ~600 credits were wasted

Silent clips first, fix lips later

Six silent Martin clips (108 credits) were thrown away, plus 131 credits on a lip-sync tool. Make the talking clip WITH the voice, once.

Voice piece started mid-sentence

Pieces cut at shot starts began inside a word. The mouth then missed the first words. Start 0.1 s before the first word.

Piece under 2 seconds

A 1.7 s line was refused (422). Pad with silence to 3 s.

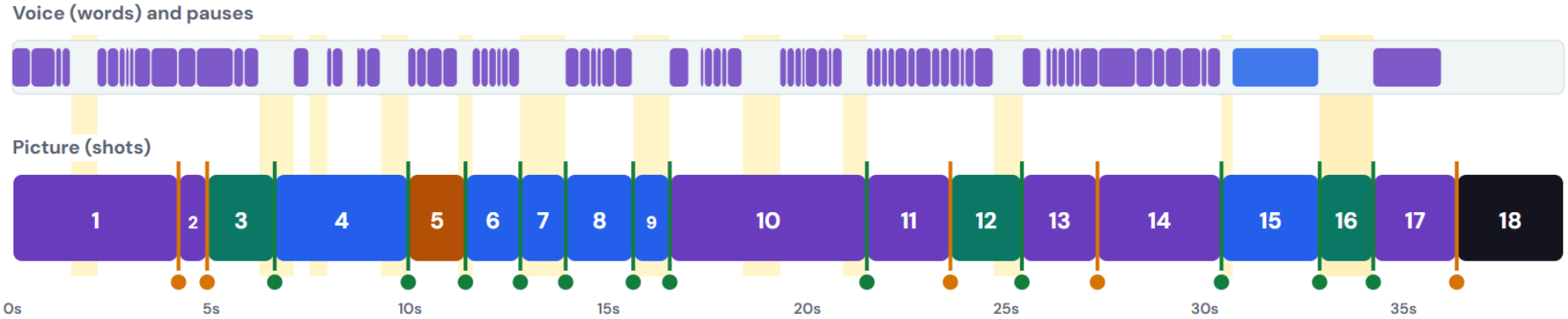
All eight clips at once

Gaurav asked for one at a time. Make one, check it, then the rest. A bad prompt is then one bad clip, not eight.

Cut rules on the real timeline

The cut rules, shown on the real Part 2 timeline

Every shot change should sit in a pause of the voice. Green marks = cut inside a pause (12 of 17). Amber = cut on a word boundary with no pause (5: a known weakness, to fix in Part 3).



Shots (numbers are on the picture bar)

- 1 Martin + Ridhima two-shot (0 to 4.18 s)
- 2 Martin (4.18 to 4.9 s)
- 3 Ridhima side (4.9 to 6.6 s)
- 4 Typing "Make him stop." (6.6 to 9.96 s)
- 5 X snaps (9.96 to 11.4 s)
- 6 Xtrusio AI (11.4 to 12.78 s)
- 7 Calling (12.78 to 13.92 s)
- 8 Connected (13.92 to 15.62 s)
- 9 Gaurav screen (15.62 to 16.54 s)
- 10 Martin, fed up (16.54 to 21.5 s)
- 11 Martin, proud (21.5 to 23.6 s)
- 12 Ridhima at screen (23.6 to 25.4 s)
- 13 Martin, smug (25.4 to 27.3 s)
- 14 Martin (27.3 to 30.42 s)
- 15 Gaurav screen: G speaks (30.42 to 32.89 s)
- 16 Ridhima, silent (32.89 to 34.24 s)
- 17 Martin: who said that? (34.24 to 36.34 s)
- 18 End card (36.34 to 39.04 s)

Martin talking (lip-synced) · Ridhima, only where approved · Screens (drawn in code) · X insert · End card (designer)

a pause in the voice (0.35 s or more) · Martin's words · G's line

Martin stops after "a team"; G follows at once; the music runs to the end (not shown).

Credits and time: Part 2 actual and the target for Part 3

Higgsfield credits only (ElevenLabs and Flow are on other balances). Balance went 1,863 to 1,282 in the picture-to-cut stretch, plus 18 for the X clip made earlier.

Part 2 actual: about 599 credits



Useful: 268 credits **Wasted: 331 credits (about 55%)**

- Talking clips used (7 x 28): 196
- Reaction clips used (3 x 18): 54
- X wake-and-snap (1 x 18): 18
- Unused silent Martin clips (6 x 18): 108
- Sync Lipsync 3 tool (9 jobs): 131
- Seedance 2.0 audio test + silent two-shot: 36
- Unused Seedance 2.5 clips (2 x 28): 56

Unit prices (Higgsfield, 720p)

Seedance 2.5 omni reference, 4 s, talking	28
Seedance 2.0 standard, 4 s, silent	18
Sync Lipsync 3 (do not use)	14.6
Seedance 2.0 audio-reference test (do not use)	18
Flow pictures	free
Screens, end card, fit, foley, mix	free

Target for Part 3 (same length, about 40 s)

9 talking clips x 28 = 252, plus 4 silent clips x 18 = 72, plus 10% for one retry = about 350 credits (against about 600). Time target: 2 to 3 hours from locked voice to check video. Part 2 took about 4 hours from the first lip-sync clip to the last cut (the hours spent earlier on script, voices, pictures were not logged). Part 3 will log every stage: start, end, credits, notes (see rules/COST-TIME-LEDGER.md).

Optional: camera move first

Optional module: plan the camera move first (from the Instagram reel we studied)

The reel was one unbroken 30 s shot. Its caption says the move was planned in a 3D tool with a plain dummy figure, and then the AI "only had to follow". NOT TESTED by us yet.

1 Plan the move

A simple 3D scene with the set and a plain figure. Put the camera path in: orbit, pull-back, push-in.

Tool: Higgsfield Scene Builder 3D (to test)

2 Render the dummy

Export a rough video of the move. No faces, no detail, only where the camera goes.

Tool: Scene Builder 3D render

3 AI follows it

Give the dummy video to the video model as a motion or video reference, with the character picture.

Tool: Seedance 2.5 video_references or Genjutsu motion control (to test)

4 Show the proof

Split screen: the finished film on top, the plain dummy move underneath, in step.

Tool: ffmpeg / Python

5 Loop the ending

Finish on the opening frame so the reel replays with no seam.

Tool: plan it in step 1

What we saw in the reel (measured)

No cut in 300 frames. Opens on a centred face looking at the lens with no title. No text burned into the picture: the story sits in the caption (hook line, problem, "so I built the move first", call to action). Use it for mood and product reels, not for talking scenes like Part 2, which need cuts in the pauses.

What we learned from ReelShort

What we learned from ReelShort (and the other reels we studied)

Two ReelShort episodes (Surf God, Triplets: 130 shots) plus an Instagram drama made in Dreamina (37 shots). Only what held in more than one source became a rule.

Pace

- A new shot every 1 to 2 seconds: Triplets 1.37 s on average (99 shots in 135 s), Surf God 2.24 s (31 shots in 69 s).
- The slower one uses slow push-ins inside a shot instead of new angles.
- Part 2: 18 shots in 39 s, about 2.2 s each.

Framing

- 55 to 71% of shots are tight. Faces fill 60 to 80% of the frame, background soft.
- The room is shown ONCE, wide, at the start of a beat, then we live in close shots.
- About a third of shots are over-the-shoulder: the other person's hair or sleeve is blurred at the edge.

Continuity

- One fixed left/right geometry per scene: each person keeps their screen side and gaze side.
- One light direction, one colour anchor, one costume per person.
- Close shots have a plain soft background, so nothing can contradict the wide. Re-use a framing and cut back to it (10 to 32% of shots).

Camera and cutting

- Change the shot size every shot: wide, extreme close-up, insert, over-the-shoulder, low angle.
- A foreground object (veil, plant, partition edge) wipes across the lens between shots.
- Extreme close-ups of ONE feature (eye, mouth, hands) carry emotion in about a second. Cuts follow a look or a gesture. Pushes run while the line is spoken.

Sound

- The speaker never stops: the picture cuts over his voice, the voice keeps running.
- Music from second 0, no quiet intro (measure it). Foley LAST, on what the finished picture shows.
- When a second voice interrupts: the first drops about 8 dB and keeps going, the interrupter rises about 6 dB, silence only once at the payoff.

Text and ending

- ReelShort burns names and spoken lines into the picture as captions.
- It ends on a held reaction close-up plus one physical cue.
- The cause of the plot must be visible on screen (in Part 2: the line typed and the call connecting).

Part 2 against these rules

Followed: faces in tight close-ups, the opening over-the-shoulder two-shot, a held reaction at the end, music from second 0, the speaker never stops, foley last.

Not yet: burned-in captions (none in Part 2: a decision for Gaurav), foreground wipe objects between shots, extreme close-ups of eye, mouth and hands as inserts (only one so far: the reflection in his glasses).

Architecture v2: the whole line (3 Nov 2026)

Pictures: `architecture/01-the-whole-line.png` (open `architecture/index.html` to see all six). This file is the same thing in words. If the process changes, edit `architecture/build_architecture.py` (the DATA blocks) and run it again, so the pictures never go out of date.

Merged from three sources: Part 2 of the Xtrusio office series (what we actually did, with the waste removed), the Base360 ad playbook (`references/BASE360-PLAYBOOK.md`: sound first, gates, no credit before the voice is locked) and the ReelShort study (`rules/REEL-STYLE-RULES.md`, `architecture/06-...png`).

Three phases, 13 stages, 7 gates

PHASE 1 SOUND FIRST (no Higgsfield credit until G2, G3 and G4 are passed) | # | Stage | Tool | Gate | |---|---|---|---| | 0 | Brief and script: one-line idea, word-for-word script, who speaks, who only listens, how it ends | Claude + Gaurav | G1 Gaurav accepts the words | | 1 | Voices: one ElevenLabs voice per character, Eleven v3, emotion tags, whole speech in one take per speaker, lock the file | ElevenLabs TTS | G2 Gaurav listens | | 2 | Pause map: transcribe the locked voice with word times, list every pause over 0.35 s | faster-whisper | none (data) | | 3 | Music: one cue longer than the reel, audible from second 0, swell near the end, never cut | ElevenLabs Music | with G2 |

PHASE 2 PICTURES AND CLIPS (one at a time, check the first) | # | Stage | Tool | Gate | |---|---|---|---| | 4 | Shot plan: one shot per sentence, speaker in tight close-up, cuts only in pauses, listener only where Gaurav says, open on a two-shot never a top view | Claude | G3 Gaurav approves the shot list | | 5 | Pictures: one per shot, character reference pasted into EVERY prompt | Google Flow | G4 Gaurav approves pictures | | 6 | Screens (phones, monitors): drawn in code, warped onto the picture | Python | seen at G6 | | 7 | Talking clips: Seedance 2.5 omni_reference, picture + that sentence's voice piece. ONE clip first | Higgsfield, 28 credits per 4 s | G5 first clip approved | | 8 | Silent clips (reactions, inserts): Seedance 2.0, 4 s, one at a time | Higgsfield, 18 credits per 4 s | yes before each paid clip |

PHASE 3 FIT AND FINISH (mostly free) | # | Stage | Tool | Gate | |---|---|---|---| | 9 | Fit to the voice: find the same words in the clip, set offset and speed | `templates/align_clip_to_voice.py` | none | | 10 | Check video: everything on the locked voice and music, every cut in a pause | `templates/build_animatic_part2.py` | G6 Gaurav watches the reel | | 11 | Foley LAST on real actions found in the frames; end card designed by the designer | `templates/build_part2_mix_v5.py` | seen at G6 | | 12 | Final check by a fresh reviewer, then Gaurav's yes. Nothing is posted without it | reviewer agent | G7 Gaurav says yes |

Why this order

- Sound first (Base360 rule): if the emotion is in a locked, continuous sound, a small picture mistake passes. And no paid picture is made on words that may still change.
- One talking clip per SENTENCE, made WITH its voice (Seedance 2.5 omni_reference). This is the same format as the Base360 ad. Sync Lipsync 3 and "silent clip first, lips later" are NOT used (Part 2 wasted about 240 credits on them).
- The word-anchored fit stays, because the model re-speaks the line a little early, late, fast or slow. Base360 had the same step (`templates/base360-tools/picture_fit.py`).
- Cut only in pauses of the voice (diagram 03). The speaker never stops while the picture cuts (ReelShort).

Roles

Gaurav: words, ears, every paid step, every post. Claude Creative Director: delivery, pace, shot sizes, small camera settings, suggestions unasked. Designer agent: end card and brand look. Reviewer agent: fresh eyes at stage 12. Final CEO session: reports only, no routine relays.

Optional module

Plan the camera move first (diagram 05, `references/CAMERA-MOVE-TECHNIQUE.md`). Not tested by us. For mood or product reels, not talking scenes.

Lessons from Part 2 (Xtrusio office series, 3 Nov 2026)

Each item: what happened, then the rule. Gaurav's rules are marked (G). The 40 s reel went through 12 check videos (v1 to v12).

A. Lip-sync and talking shots (the expensive part)

1. Silent clips first, fix the lips afterwards: wrong. Six silent Martin clips (108 credits) and 9 Sync Lipsync 3 jobs (131 credits, slow serial queue, 10+ minutes each) were thrown away. RULE: make the talking clip WITH the voice: Seedance 2.5 omni_reference, picture + that sentence's voice piece (same as Base360). No Sync Lipsync tool. (G, 3 Nov)
2. Voice pieces cut at shot starts began mid-word, so the mouth missed the first words. RULE: one piece per SENTENCE, starting 0.1 s before the first word.
3. Seedance refuses audio under 2 s (422). RULE: pad with silence to 3 s.
4. All clips in one batch: wrong. RULE: ONE clip first, check it, then the rest, and only on Gaurav's word. (G)
5. The model re-speaks the line a little early/late/fast/slow. RULE: word-anchored fit with `templates/align_clip_to_voice.py` (clip tag `clip:<name>@<offset>*<speed>`). Fewer than 3 matched words = remake the clip.
6. A mouth-opening score (mediapipe) did not agree with what the ear heard (it follows loudness). RULE: judge lips with eyes on frames and with Gaurav, not with that score.

B. Editing rules from Gaurav's notes

1. Cut only in pauses; never in the middle of a sentence. Still open in v12: "viral on | Instagram" and "clap | while he plays director". (G)
2. Ridhima appears only where approved. List the placements and get a yes before dropping her in. She must show in the opening for context (over-the-shoulder two-shot), and in side pose at "Instagram", then she types. (G)
3. No top or wide view on talking lines; two close-ups, faces on talking. (G)
4. After "a team", go straight to the Gaurav screen. No more Martin dialogue. Martin stops, G follows at once. (G)
5. Music runs from second 0 to the very last frame, building up. Start the scene on a sound. (G)
6. Screen text must be readable: popup on the LEFT side when Ridhima is in shot. Zoom into the screen and place the real UI (drawn in code). (G)
7. Do not remake a clip that was fine: trim and time it. (G)
8. The cause of the plot must be visible (the typed line, the call connecting).

C. Foley and sound

1. Foley LAST, on actions found in the finished frames: typing keys on the picture, snap on the snap frame, ping, ting/tung, chime, desk thump at 0.05 s, X snore on the end card.
2. A real snap sound came from an ElevenLabs sound effect (`sfx2/snap-1.wav`), placed on its measured peak (52 ms in).
3. Music cue 15.7 to 54.74 s of the 40-second-of-fear file, fade only the last 0.6 s.

D. End card

1. My own versions were rejected (blue, thin plain font, wrong order). RULE: the DESIGNER makes it, in the brand look: black full screen, small white text, DM Sans, designed like a mobile app screen, "Follow for Part 2". No extra copy lines. Template: `templates/endcard-follow/`. (G)

E. Flow and Higgsfield facts

1. Flow: paste the reference into EVERY prompt or the face drifts; 2 images per prompt; Enter sometimes fails, click the send arrow. A frozen Flow tab: close it and reopen.
2. Higgsfield uploads: copy the long presigned URL exactly (a typo gives 400); one upload per Bash call; `media_confirm` after the PUT.
3. A preset recommendation blocks the call: retry with `declined_preset_id`.
4. `jobs_wait` returns at once: poll with timed waits.

F. Working with Gaurav

1. Give suggestions unasked, and decide cosmetic things yourself. (G)
2. Say what you understood before big moves. (G, Base360 rule 22)
3. Show credits and get a yes before any paid step; keep a running balance.
4. Reply in simple English, short.

G. Tool limits found

1. Instagram blocks fetches to localhost (CSP) and blocks autoplay in the Chrome tab: reel studies are done by reading frames inside the page, and sound could not be measured.

Cost and time ledger

Higgsfield credits only (ElevenLabs and Flow have their own balances, noted separately).

Unit prices (720p, 9:16)

Item	Credits
Seedance 2.5 omni_reference, 4 s, talking (7 per second)	28
Seedance 2.0 standard, 4 s, silent	18
Sync Lipsync 3 (do not use)	about 14.6
Flow pictures, screens, end card, fit, foley, mix	free

Part 2 actual (about 40 s reel, 18 shots)

Total about 599 credits: useful 268 (7 talking clips 196, 3 reaction clips 54, X wake-and-snap 18), wasted 331 (6 unused silent Martin clips 108, Sync Lipsync 3 131, Seedance 2.0 audio test + silent two-shot 36, 2 unused Seedance 2.5 clips 56). Balance went 1,863 to 1,282 over the picture-to-cut stretch, plus 18 for the X clip earlier. Time: about 4 hours from the first lip-sync clip to the last cut. Script, voices and pictures were not logged.

Part 3 target

9 talking clips x 28 = 252, plus 4 silent x 18 = 72, plus 10% for one retry: about 350 credits. Time: 2 to 3 hours from locked voice to check video.

Part 3 log (fill in as you go; Part 3 is the measured test)

Stage	Start	End	Minutes	Credits (start to end balance)	Notes
0 Brief and script				free	
1 Voices				ElevenLabs:	
2 Pause map				free	
3 Music				ElevenLabs:	
4 Shot plan				free	
5 Pictures				free	
6 Screens				free	
7 Talking clips					first clip checked at:
8 Silent clips					

Stage	Start	End	Minutes	Credits (start to end balance)	Notes
9 Fit				free	
10 Check video				free	
11 Foley + end card				free	
12 Final check				free	
TOTAL					against 350 and 2 to 3 h
Also record: number of check-video versions (Part 2: 12), clips remade, and every Gaurav note that came after G6.					

Reel style rules (Gaurav, 3 November 2026)

These apply to every reel and series from now on. They come from studying real short dramas and from Episode 1.

Framing: faces, emotion, hidden errors

1. **60 to 80 percent of the shots are faces.** The face fills the frame, the background is soft blur. The story is told by emotion, not by the room.
2. **Show the place once, for about 1 second, at the start.** Use a TOP angle (aerial, ceiling) or a GROUND angle (low), not a busy wide shot. Like a serial that shows a city or a building before going into the house. Small mistakes in the room hide in a top or ground view.
3. After that, go straight to faces. Dialogue is framed around the faces.
4. The listener's face is the picture while the other person talks. The speaker's voice keeps running over it.
5. A wide shot of both people is allowed once per beat, short, at the moment the story needs it (for example when the line is about both of them).
6. Insert shots (a screen, an object) only when they carry the story, and only while the voice refers to them. Screen text must be exact and sharp.

Voice and sound

1. **The speaker never stops talking.** Cuts happen over his voice (the picture changes, the voice continues). Natural pauses and emotions are written into the text with tags.
2. Words are simple. If the audience would ask "what is that?", change the word.
3. When a second voice interrupts, the first speaker KEEPS talking underneath. In the mix the first goes down about 8 dB and the interrupter up about 6 dB. Then the first speaker trails off in confusion, then silence. Silence is used once, at the payoff.
4. Music is present from second 0. No quiet intro. Check the loudness over time by measurement, not by trust.
5. Foley comes last, placed on what the finished picture shows. Never before the picture exists.

Camera

1. Almost no camera moves. A move is allowed only to show the space or to follow a look or gesture. Camera dancing is not needed.
2. Every push or zoom runs WHILE the line is spoken, from the first word to the last. Never after.
3. Cuts follow a look or a gesture. Re-use set-ups (cut back to a frame already made).

Continuity (from the reel studies, see `references/CONTINUITY-RULES.md`)

1. One fixed geometry per scene: each person keeps their screen side and gaze side.
2. One light and one colour anchor per scene. One costume per person per scene.
3. Props repeat as anchors (a mug, a plant, a sign).
4. End on a held reaction close-up plus one physical cue (for example a reflection in the glasses, a snore).

Story

1. The cause of the plot must be on screen: the viewer has to SEE what starts the chain (for example the line typed on her screen and the call connecting).
2. If the listener has no lines, her conflict comes from silence and small actions.
3. Keep the length honest: the voice sets the length. A longer speech means a longer scene, so re-time the pictures.