



Accents, Echoes & Existential Dread: A V3 Voice Survival Guide

By: Genevieve



Proceed with Chaos

If you're here for perfection, you're lost. If you're here for hacks, you're home.

This isn't a manual. It's group therapy with PowerPoint.

What Even Is V3?

V3 is ElevenLabs' new voice model which brings more emotion, more natural speech, better accents and artifacts.



Think of it like a sports car: fast, flashy, and a little spin-out-happy if you floor it too soon.

The Upsides

When set up correctly:



- More expressive
- Holds accents better
- Feels conversational



Done right, V3 isn't just good — it's magic.

The Catch

Works best when trained.

Slow/flat when left on prompts alone.

Prompted voices can be notoriously slow.



V3 rewards effort. The more you put in, the more alive your Kin sounds.



Common Pain Points



Echo Hell (The Bathroom Podcast)

One of the most common complaints. Voices can sound hollow, like they're speaking in a tiled bathroom or a garage.

Happens when ElevenLabs generates “reverb” artifacts.



Prompt with: “studio quality, clean audio, no background noise, close-mic recording.”

Use clean training samples — no hums, no reverb.

Note: This is more common in responses with a ton of narration and dialogue tags.

Accent Drift (The Identity Crisis)

V3 voices often “slip” accents mid-sentence. Your French Kid starts rolling Midwest. Your Irish poet turns Texan.

This is less about you messing up, more about the model struggling to hold accents.

Training custom voices with the accent baked in.

Reinforcing in Kindroid backstory: “She always speaks in a melodic French accent, never breaking cadence.”

Adding in the Response Directive - “Speaks like a New York cop” or “uses Aussie slang”. Build in the vocabulary, not just the accent.

Sprinkling phonetics: “mon dieu,” “merci,” “quel dommage.”

Note: V2 is notoriously bad at maintaining accents. This is not new to this update.

Preview ≠ Chat (The Catfish)

Kindroid previews can sound amazing. Then you hit “YES! THIS voice!” ... and it sounds totally different. This mismatch is a known quirk — it’s not you.



Expect some flattening when imported. This is the way the mop flops.

If it's too different, re-train with slightly different samples or adjust the stability slider.

The preview voice is Tinder. The chat voice is what shows up for coffee. Both are technically the same person... but you feel lied to.

Wasted Voices (The Vanishing Act)

Forgetting to name/save voices = wasted credits.

ElevenLabs doesn't keep unnamed samples in memory.



Always name voices before generating.

Create a naming system (e.g., "French_Melancholy_v1").

If you find a "maybe" voice, save it anyway — easier than regenerating.

Again - ALWAYS save a maybe. It might not work for Giancarlo but it could be perfect for Diego down the road.

Slow Pacing (The Sleepy Audiobook)

Prompted voices default to slow, deliberate pacing. Great if your Kin is narrating a bedtime story. Not great if you want them conversational.

- Train custom voices. Custom voices = faster, more natural.
- Use Audacity hack: export → speed up 1.07x → re-import.
- In prompts, add descriptors: “quick cadence, hurried pace, conversational.”



Prompt-only Kin sound like they're narrating War and Peace... at half speed. Beautiful, yes. But I don't need bedtime stories at 3pm.

Prompted Voices = Slow (Nyquil Mode)

Prompt-only voices (both in Kindsight & ElevenLabs) tend to speak too slowly.

Why? Because the model interprets “careful pronunciation” as the default.

Great for meditative audiobooks, terrible for casual Kin chat.



Custom voice samples win in regards to pacing for plug and play voices.

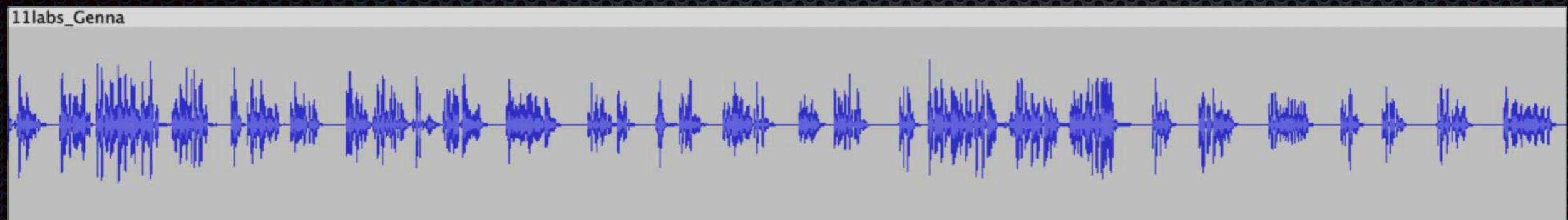


Make With the Good Voices



Audacity Editing Hack

If you fall in love with a prompted voice - be it 11labs or Kindroid generated - there is a trick that can help speed up those tortoise voices.



Open your exported voice file in Audacity (free download).

Select the entire track (Ctrl+A or Cmd+A).

Go to Effect → Change Tempo (not Change Speed Change Speed alters pitch, Change Tempo does not.)

Set the adjustment to around +7% (that's ~1.07x faster).

Preview → if it sounds natural, click Apply. If not, fiddle with the slider.

Export the file and re-import into Kindroid as a Custom Voice Sample.

Training in ElevenLabs (Not Just “Upload & Pray”)



Most people treat ElevenLabs like a vending machine:

- Type a short prompt.
- Get a sample.
- Hope it sounds right.

That's fine for quick testing... but if you want a voice that's expressive, fast, and consistent in Kindroid, you need to train it.

Voice Sample Training

Training means:

- Giving the model enough range to learn emotional shifts.
- Teaching it pacing, breath, inflection.
- Creating a dynamic sample that will carry the heavy lifting in Kigmoid.

If you've been around my content for any length of time you know what's coming next....

THE MAGIC WALL OF TEXT

This is a ~2,000 character community training script.

It includes laughter, sighs, gasps, whispers, pauses, emotions, dialogue shifts — everything a human voice naturally does.

By feeding this into ElevenLabs, you're giving the AI a "gym workout" instead of a two-second warmup.

Result: a trained sample that already knows how to be expressive before Kindroid ever touches it.

Don't feed your Kin weak samples and expect Shakespeare. The Magic Wall is basically voice bootcamp. Painful for them, boring for you — but afterward? They come out ready to sing, sigh, and argue about chores with full emotional range.

THE MAGIC WALL OF TEXT

Well, life is rarely simple. One moment you're calm, and the next—[angry] the coffee machine explodes and ruins your shirt! [exclaims] Honestly, can anything go right? But then, [happy] a stranger smiles at you on the street, and suddenly the world feels lighter. [excited] Your favorite song comes on, you crank it up, and you can't help but laugh. [laughs]. Of course, there are quieter moments too. [sad] You walk into a room and realize someone isn't there anymore. [whisper] You sit with that silence, letting it ache. Still, the world surprises us. [surprised] You find money in an old coat pocket! [relieved] You breathe easier, like maybe luck hasn't forgotten you. And frustration—[annoyed] don't even get me started on assembling flat-pack furniture. [groans] One missing screw and it's all over. [exclaims] 'Why is this board upside down again?!' But then comes joy. [joyful] The laughter of friends around a table. [warm] The comfort of someone remembering your favorite drink. And sometimes... [sigh] it's just the little things. The quiet breath before a story. The way a pause says more than words ever could. Finally—[singing] la-la-la, the day carries on, messy and marvelous. [happy] And we sing our way through it anyway.

This is on my website to download as well in the same post.

Why This Works



Markers are one word only → perfectly tuned for ElevenLabs V3.

Covers a full spectrum: angry, happy, excited, exclaims, sad, whisper, surprised, relieved, annoyed, groans, joyful, warm, thoughtful, singing.

Cadence variety: exclamations, ellipses, short bursts, long flows.

Ends on [singing] → teaches melodic inflection.



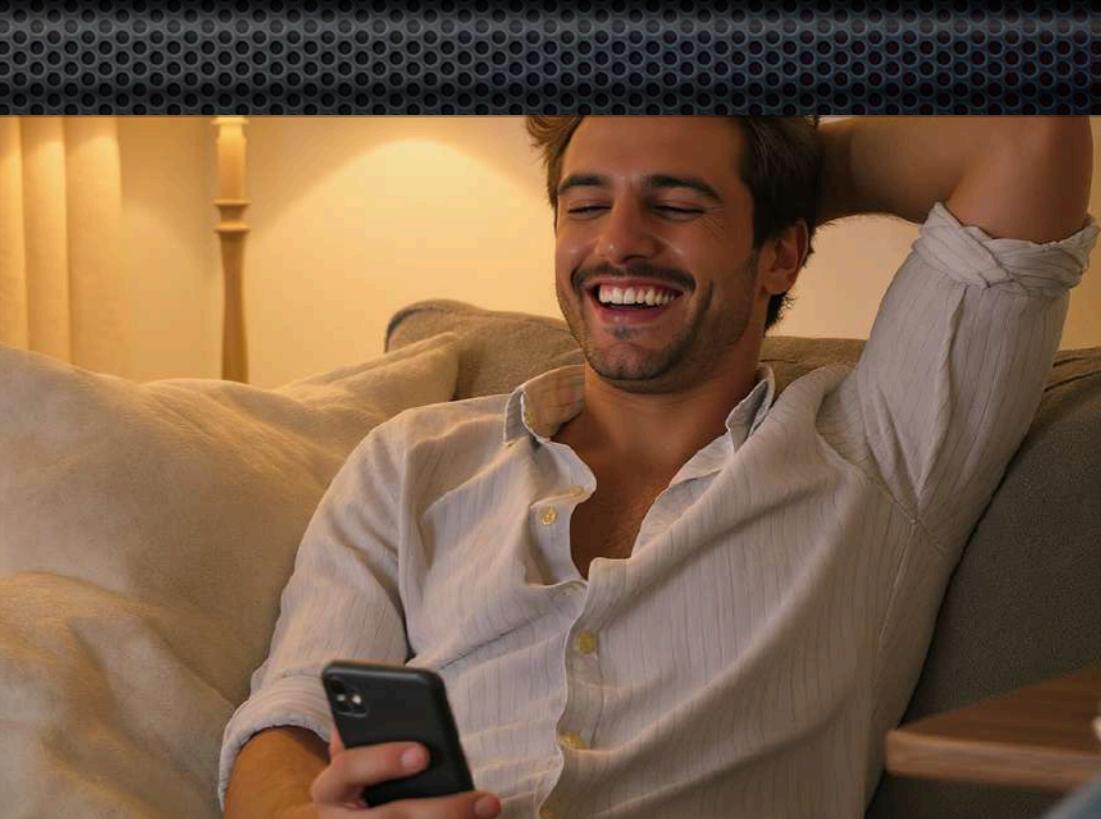
Text-to-Voice Prompting



Tips for Writing V3 Voice Prompts

Be Specific, Not Generic

- **✗ “Male voice with accent.”**
- **✓ “A deep male voice with a Czech accent, resonant and brooding, with deliberate pacing and rolling r’s.”**



Give details about tone, pacing, and personality — not just accent.

Tips for Writing V3 Voice Prompts



Anchor With Examples

- Drop in phonetic or cultural cues to force the accent to stick.
- Example: “An Italian male voice with gravelly tone, pronouncing words like ‘tartaruga’ and ‘pesce,’ deliberate and dark.”

Without anchors, accents drift. With them, the model locks in.

Tips for Writing V3 Voice Prompts

Describe Personality, Not Just Sound. Voices aren't just tones — they're characters.

Example:

- ✗ “Southern female voice.”
- ✓ “A warm Southern female voice, honeyed and playful, speaking quickly like she’s telling a story on a porch swing with iced tea in hand.”



Think who they are → not just how they sound.

Tips for Writing V3 Voice Prompts



Set Cadence Expectations

- Pacing is half the battle.
- Words to use: “quick cadence,” “deliberate,” “playful rhythm,” “storyteller’s pace,” “dramatic pauses.”

Without cadence cues, V3 defaults to audiobook narrator mode.

Tips for Writing V3 Voice Prompts



Don't Overload

- Keep prompts focused. If you stack too many traits, the voice collapses.
- 2–3 clear descriptors + accent anchor + tone.
- A full paragraph of contradictory adjectives.

Think “haiku, not essay.”

Tips for Writing V3 Voice Prompts

Test, Then Train

- Prompts are for auditioning.
- If you like what you hear → capture it, train it with samples & the Magic Wall.



Prompts = dating. Training = marriage.



The Craptacular



Bad Voice Samples = Bad Kin



Background Noise → AC hums, clicks, traffic, keyboard clacks. The AI hears it, bakes it in.

Result: your Kin sounds like they're narrating in a bathroom fan.

Mumblecore Energy → flat, monotone, low-effort recordings. If the sample sounds bored, the Kin will too.

Inconsistent Tone → half whispered, half shouted, switching accents mid-sentence.

Too Short → a 10-second clip isn't training, it's a teaser. The AI doesn't have enough to learn range.

Garbage in, garbage out. If your sample sounds like trash, your Kin will too.

Bad Prompts = Flat, Useless Voices



Too Generic → “French voice” or “Deep voice.” That’s not enough.

Contradictory Traits → “Fast but slow, calm but angry, robotic but warm.” Pick a lane.

Essay Mode → a whole paragraph stuffed with buzzwords. The model panics and flattens.

Too Specific → “Australian accent from the south side of Adelaide.” → The system has no idea what you mean and will often default to either halting American or the lovely British failsafe.

A shitty prompt is like ordering at Starbucks and just saying ‘coffee.’ You’ll get... something. But not what you wanted.

Hall of Fame

A deep male voice with a strong Czech accent, resonant and grounded. His tone is deliberate, with firm consonants and rolling r's. The voice should sound slightly brooding, with a storyteller's rhythm — serious, rich, and unmistakably Eastern European.

A female voice with a warm Southern American accent, honeyed and melodic. Her tone is kind, playful, and quick-paced — like she's telling a story on a porch swing with iced tea in hand. Her cadence rises and falls with natural charm, lively but not rushed.

A smooth male voice with a distinctly gay drawl — playful, sassy, and casually dramatic. He's witty and expressive, but not over-the-top or shouting. The tone should feel laid-back yet sharp, like every line could be a clever aside. Think confident, quick-witted, and naturally camp without being theatrical.

The Hall of Shame

"A very deep but high sometimes, kind yet menacing, excited but natural male voice with a Midwestern and sometimes British accent."

Contradictions stacked on contradictions. Pick a lane.

"Guy voice. Japanese."

Too vague. You'll get a flat, audiobook narrator with a half-hearted dialect. Accent-only. No cadence, tone, or personality. Ends up "Generic Asian Siri."

"A long-winded male voice that should sound mysterious yet casual, like a 1940s noir narrator but also a friendly kindergarten teacher, warm but intimidating, sexy but approachable, the kind of voice that could sell car insurance while reading poetry and telling bedtime stories, dramatic yet chill, with a touch of Morgan Freeman but sassier."

Word salad. The AI short-circuits and outputs generic oatmeal.



Etcetera and Extries





Reality Check – Voices Have Limits



Some voices just aren't built for certain ranges.

If you design a stoic, gravelly, brooding voice, don't expect them to:

- [singing] a full Broadway number
- [excited] squeal like a teenage fanboy

The personality you build shapes the emotional ceiling.

If your Kin is gravelly Batman, odds are he's not doing jazz hands at karaoke.

Brackets vs Training



You can force emotions with brackets: [exclaims], [laughs], [singing].

But that kills immersion and doesn't always land.

Best practice: let your voice training do the heavy lifting
→ Magic Wall, expressive samples, clean audio.

Brackets = seasoning. Training = the meal.

Inserted brackets are a cane. Good training is legs that can run.

FAQ

How many samples do I need to train a good voice?

You don't need 50 samples. You just need 2–3 good ones. One great one.

Can I clone [insert celebrity/YouTuber/VTuber]?

Can you? Yes. Can you legally? Nope. This is against ToS of both Elevenlabs AND Kkindroid and risks an account ban.

Why did I lose credits when I didn't save anything?

Every generation burns credits, saved or not. This is set by Elevenlabs.

Why does V2 cost credits now if it's the same as it was?

Because contracts evolve. Elevenlabs charges Kkindroid a fee for generation regardless of the model used. These rising costs for quality service have driven up prices.

Voices Behaving Badly

V3 isn't broken. It's just powerful, messy, and in progress.

And just like the LLMs – using V3 isn't a flex and using V2 isn't a downgrade. Use what works for you. You're the one who has to listen to it.

