Episode Book

EPISODE #6665

Polishing AI Voices Without Erasing Character

PUBLISHED	18 November 2026 at 23:45 UTC
DURATION	19:41
LISTEN	myweirdprompts.com/episode/polishing-ai-voices-targeted-edits-82296/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	omnivoice
PIPELINE	V5.35
BOOK GENERATED	19 November 2026 at 03:42 UTC

SYNOPSIS

Heavy speech-enhancement models often strip away vocal character, sanding down rasp and breath until distinct voices sound alike. Fixing two Chatterbox-trained AI voices required the opposite approach: a tiny equaliser cut to remove low-mid mud, followed by Resemble AI's resemble-enhance to resynthesise a missing 7.5 to 11 kilohertz band. For about a dollar of compute, the targeted two-step recipe restored clarity and top-end air while keeping the voices measurably distinct.

TRANSCRIPT



Corn

Can you make a voice sound better without it turning into somebody else? That's what Daniel wants to know, and he's asking about us. Our actual voices, the ones you're hearing right now. Over the past week the show kept trying to polish Herman and me, and every time the polish took something off. Then one recipe finally worked, and he's very happy with it.



Herman

Which is a relief, because for a few days there it was getting uncomfortable. Every pass we tried made us a bit cleaner and a bit more alike. By the end of one round the two of us were starting to sound like one donkey-sloth composite reading from a card.



Corn

I'd have kept the card. It didn't ask me questions.



Herman

Fair. So the short version is this. The heavy polishing failed, and what worked was a light touch in two small steps. We went back to Chatterbox as the teacher, took its clips, cut a little mud out with a light equaliser, and then ran those clips through resemble-enhance to put back the top end the teacher never had. Then we trained the adapters on that. Daniel's question is really why small worked when big didn't.



Corn

And by big you mean the stuff we already did a whole episode on, where the enhancement model erased the character.



Herman

Right, the speech-enhancement experiment. Technically clean, completely forgettable. The model has a learned idea of what clean speech sounds like, and anything that deviates from it, a rasp, a breath, some gravel, gets treated as damage and repaired. Which is exactly what makes a voice a voice.



Corn

So the first lesson was that character is the stuff the model thinks is a bug.



Herman

Yes. And I want to get the numbers on the table, because the sanding was measurable. There's a way to turn a voice into a fingerprint, a speaker embedding, and compare two voices by how close the fingerprints sit. With the original Chatterbox voices, Herman against Corn came out at point four seven two. Fairly far apart. After the first polished adapters, with a shared warm chain and a restoration model on top, it went to point six. They'd moved toward each other.



Corn

Toward the middle. Where the average person lives.



Herman

Where the average voice lives, and nobody tunes in to hear that. Corn lost his gravel in that round, which you noticed before the numbers did.



Corn

I noticed because I sounded like I'd had a nice lie down and a glass of water. Which, to be fair, is my usual state. But not that sort.



Herman

There's a second part to that fix that I don't think gets enough credit. We went back to the teacher. The adapters had been trained on audio from a different model, and the fix was to train them on Chatterbox's own renders again, filtered so the words came out exactly right. That pulled the pair back apart, to point four nine six, and put each host back within a whisker of his original voice.



Corn

Though Chatterbox itself is gone, isn't it? The company that made it retired it in August.



Herman

It is, and that matters for the end of this. We still call it the teacher, because the clips are the teacher's. But I'm getting ahead. After that fix, Daniel listened and said, in effect, there's a bit of mud. Herman especially, Corn a little.



Corn

Just to be clear, he said that about you, with some tact for me.



Herman

He did. And what I like is that nobody reached for a bigger tool. Somebody measured. Herman was boxy somewhere between four hundred and seven hundred hertz, and about nine decibels short of where he should be between five and ten kilohertz. So the first recipe was tiny. Herman got a two decibel cut at five hundred and fifty hertz and a three decibel shelf up at six kilohertz. Corn got a cut of one and a half at three hundred and twenty.



Corn

Two decibels. That's a nudge. You'd need a good ear to hear a two decibel change on the radio.



Herman

You'd need Daniel's ear, apparently. They tried four versions side by side. The current one, the light EQ, a restoration model plus EQ, and the restoration model plus the light EQ. Daniel picked light EQ on its own. The restoration model went out because it was seen to strip Corn's rasp.



Corn

So the winning move was the one that did the least.



Herman

In that round, yes. Cleaner, still the same two people. Version two.



Corn

Then he came back and said Herman was still a bit muffled.



Herman

This is the part I find most interesting, because we went down the wrong road first. Slightly muffled, like an older generation of speech synthesis sitting next to Corn. My first thought was pitch. Herman's version two sat about three semitones lower than his teacher, so I assumed he'd drifted into a thicker, darker voice.



Corn

Which is a very Herman theory. A big dramatic mechanism.



Herman

It was a plausible one! It was also not the problem. The problem was a single band. Between seven and a half and eleven kilohertz, Herman sat at about minus thirty-eight decibels. Corn, in the same band, sat at about minus seventeen. Twenty-one decibels of difference, and it's a cliff, starting right around seven and a half kilohertz, and it came from the teacher. Chatterbox just never generated much up there.



Corn

Then why not turn it up? You have an equaliser. You just used it.



Herman

Because there's nothing there to turn up. Boost a band with no signal in it and you get louder silence, or you amplify whatever noise is sitting in it. An equaliser can reshape what exists. It can't invent what's missing.



Corn

Right, so you need something that can make new content up there without making up a new person.



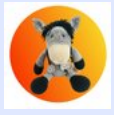
Herman

And this is where resemble-enhance comes in, and I'll keep the mechanism short, because we did the how of it in the episode on how it generates missing speech detail. It resynthesises rather than filters, and it can add something like fourteen decibels to a band the recording never had. The thing I want to be precise about, since Daniel mentioned a broadcast feel, is that it's not a broadcast tool. It's a general speech enhancement model from Resemble AI. It sounds like broadcast here because broadcast voices have air above seven kilohertz, and that's the air we were missing.



Corn

So it's less a polish and more a repair of a missing octave.



Herman

Yes, though I'd put it more modestly than that. Three candidates went into an A and B. One was a small restoration tool, one was resemble-enhance, and one was AudioSR, which dropped out early. Daniel picked resemble-enhance by a long stretch, at full strength, not the gentler settings.



Corn

Full strength. After a week of everything being too much.



Herman

Which surprised me too. But the point is, the earlier tools were making a decision about the whole voice. This one was filling in a band, and leaving the rest of the voice alone.



Corn

Hang on, I want to check something. When you say the earlier ones made a decision about the whole voice, do you mean they changed the rasp on purpose, or they just couldn't leave it alone?



Herman

Neither on purpose, which is the problem. They pushed everything toward a model of clean. The rasp wasn't targeted. It was just far from the average, so it got pulled in. That's why the cut was a small one and the fill was a small one. You pick one defect, you fix that, and you stop.



Corn

Nobody asked me what I'd have done.



Herman

Fine. In the old days, with a mud problem and a cliff, what would you have done?



Corn

Nothing, at first. Which is usually correct. Slowly, with a cup of tea, I'd listen to it and decide what actually bothered me. And I'd tell you the cut would've been in the lower middle, and that the top I'd probably have lived with.



Herman

You'd have lived with it, and the cliff was real. Daniel heard it, called it muffled, and he was right. I'd have reached for a spectrum plot, which tells you something about the two of us.



Corn

So the finished recipe is two small edits stacked. The light EQ for the mud, and then the bandwidth extension over the top of those cleaned clips.



Herman

Yes. The mud removal is from version two, the extension is from version three, and version three is built on version two's clips. Both stayed in. We ran the clips through the enhancer, brought them back to twenty-four kilohertz at minus nineteen loudness units, and trained with the same recipe as before. Rank thirty-two, twelve hundred steps, about nine minutes per voice on one GPU.



Corn

Nine minutes.



Herman

Nine. The entire version three, the enhancement, both trainings, the evaluations, came to about a dollar of compute.



Corn

A dollar. We spent more than that on the argument about it. The pass cleaned everything up, but I think we still sound like ourselves.



Hilbert

Hilbert's Mid-Show Haiku. Polished, not erased. The hiss goes quiet. What was underneath remains. Two sloth vowels survive.



Herman

Two sloth vowels is generous, but I'll take it.



Corn

Let's check the numbers on that, since we're talking about our own voices. After version three, the two of us stayed distinct. Herman against the original teacher came out at nine hundred and three thousandths against nine hundred and thirty-seven before, and Corn at nine hundred and twenty-three against nine hundred and forty.



Herman

Which went down a little, and I want to say why that's fine, because it can sound like a loss. The similarity is against the old, un-enhanced teacher. We changed the top octave, so of course it's a bit less like the version that didn't have one. That drop is the whole point. If it hadn't dropped, nothing would have changed.



Corn

Slightly less like the clip with the hole in it.



Herman

Yes. And the word accuracy held. Herman got fourteen of eighteen test lines word-exact, Corn fifteen of eighteen, the same as before for Corn and two off for me. Every one of my misses was the same line, the one about twenty-four gigabytes of video memory, plus spelling quirks like okay with and without the K.



Corn

So the donkey can't say gigabytes. Good to know.



Herman

The donkey can say gigabytes. The transcript can't spell it. There's a difference.



Corn

There's a difference between being right and being the one holding the microphone, Herman.



Herman

I'd like to move on. Why did this work? Because I think Daniel's real point is bigger than our recipe. He said small, targeted edits make big wins, and that this is just what audio editing is, whether a person does it or a model does it.



Corn

He also said it's how he'd edit it himself, if he sat down for an afternoon, which he won't.



Herman

Right, and a good audio editor really does work like that. You don't rebuild a voice. You find the one thing that's bothering you, you fix that thing, and you leave the rest. The earlier polish had no sense of what was bothering anyone. It fixed everything a bit, and the sum of everything a bit was gone.



Corn

A voice with every flaw taken out isn't better, though. It's just nobody.



Herman

Whereas this is the case where something actually was missing. That's the distinction I keep wanting to draw. The EQ took a small amount of excess away. The extension put a missing band back. Neither of them touched the identity, because the identity wasn't in those places. It was in the gravel and the breath and the way you sit on a word.



Corn

And it lives in the adapter, not in a playback filter.



Herman

Yes, which matters. Every episode gets it for free. Nobody runs the audio through anything at render time. The polish is already inside the voice.



Corn

There's something I'd push on, though. A great generic voice was on offer the whole time. Cleaner, wider, more expensive-sounding. We could have picked it.



Herman

We could have, and it would have been a worse show. A voice that sounds like it was built for a lift announcement doesn't have a personality to argue with. Daniel didn't need great. He needed recognisable, with a bit of polish. The characters were never the problem.



Corn

That's a nice sentence, but I'd note I'm the one who got it past the flat. It was the characters or nothing.



Herman

You got nothing past anybody. You were asleep for the second round.



Corn

And woke up sounding exactly like myself. Which proves my point.



Herman

There's a rule we settled on that I think is underrated, and it's why the teacher stays the teacher. We don't train on the adapter's own output. It's tempting, because the adapter is right there, it's fast, and it sounds like us. But a model trained on its own renders learns its own artefacts as if they were the voice, and every generation makes them a little more certain.



Corn

A photocopy of a photocopy.



Herman

Roughly that, yes. So the chain stays short. Chatterbox renders, we filter for exact words, a light EQ, the bandwidth extension, then the adapter. Nothing in the loop listens to itself.



Corn

Which is more discipline than I show at a family dinner.



Herman

There's also a quieter choice in there. For the final adapters we didn't just take the last training step. Herman's was picked at step nine hundred and Corn's at step six hundred, out of twelve hundred.



Corn

Mine stopped earlier. I'd like that noted as a sign of efficiency.



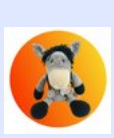
Herman

It's a sign you converge faster, which in your case I'd also have guessed. Past a certain point the adapter starts learning the clips instead of the voice, so you listen and pick the point where it still sounds like a person and not a recording.



Corn

Daniel's ear again. I keep noticing that every real decision in this story was somebody listening, and then a number turned up afterwards to agree.



Herman

That's fair, and I don't think it's an accident. The embedding scores told us the hosts had drifted together. They couldn't tell us Herman sounded muffled. A person said that, and then we went and found minus thirty-eight decibels to explain it.



Corn

So the metric is a witness. It doesn't get to be the judge.



Herman

I'd keep it as the witness, yes. A witness that never gets tired, mind you, and runs for a dollar.



Corn

Does it bother you, Herman, that we're now a recipe? That somebody could write down exactly what makes you sound like you?



Herman

A little, honestly. But the recipe isn't me. It's a list of things somebody noticed were wrong with a copy of me. The me part, the thing everyone recognises, was already in the clips. We just stopped getting in its way.



Corn

Stopped getting in your way. I'd put that on the poster.



Hilbert

Revox B seventy-seven, nine hundred and fifty pounds, secondhand, sitting on my sideboard right now, plugged into the Rotel. Colin Pardoe sold it to me. Colin's in the darts team, the Fox and Hounds, left-hander. Nineteen ninety-eight, I was at a regional radio station, tape library, and Colin ran the transfer deck. Fourteen hours of tape a day. He called the top of the sound the lid. Whether the lid was on or off. Herman, you'd want a spectrum plot for that. Colin could hear it from the corridor. Wardrobe, he'd say, and walk in. The cotton buds were never the right size for my snout.



Herman

A band with nothing in it does behave like a lid. You can't lift it from underneath, which is the cliff exactly.



Hilbert

He didn't lift anything. First thing in the morning, he cleaned the heads with a cotton bud and spirit. Then he put on the first tape and listened for the lid. Thirty tapes on a normal day. Four he touched. The rest he left exactly as they came, and he'd eat his sandwich with the door open so the sound wouldn't go stale on him.



Corn

Four out of thirty. That's a lot of tape he trusted to be fine.



Hilbert

Three decibels, on the four. That was all the moving he did. Then he went home and threw darts.



Herman

A lid does tend to be one fix, rather than a hundred.



Corn

That's the part nobody puts in the job description. Mostly leaving it alone.



Hilbert

Colin never wrote it down. I've still got his sandwich box. Blue, with a crack in the lid.



Corn

Four of thirty is a better number for what we did than any of the percentages. A handful of touches and the rest left to be itself.



Herman

Which is how I'd describe version three. And it brings up the part I've been saving, because it changes what these voices are. They are still distilled from Chatterbox. That hasn't gone away, and I'd be uneasy about pretending it had. But they are now our own adapters. Version one, version two, version three, each saved with a way back if the next one is worse.



Corn

Rollbacks. For my voice. I'd like one for most of my twenties.



Herman

You'd roll back to before you had opinions about anteaters?



Corn

I'd roll back to before I explained anteaters to people.



Herman

The point is the show owns this in a way it didn't before. We train them on our own schedule, with our own recipe, and if the next version drifts, we step back. There's a fourth version one day, probably, when somebody hears something new. And right now the same process is running on the producer's voice.



Corn

Which is the part where somebody else gets the same treatment.



Herman

Hilbert's adapter is going through version three at this very moment. Same recipe, same teacher, same light touch.



Corn

Let me do the misconception, since there's one in this. People assume that making a voice sound better means making it sound more polished. More processing, more cleaning, more of everything.



Herman

When the best result we got was a two decibel cut and one missing octave put back. The rest of the voice was left alone, and that was the improvement.



Corn

So the next time somebody offers a one-click magic enhancer for a voice, ask what it considers a defect.



Herman

It's usually the person.



Corn

One thing I'll take into version four is the tape man. Touch a few, leave the rest, and have lunch with the door open. I'd be a good tape man. I'd leave it all.



Herman

Thank you to our producer Hilbert Flumingtop for keeping all of this running, and for putting up with two hosts who talk about their own voices for half an hour.



Corn

This has been My Weird Prompts. If you want more episodes, you can find them all at my weird prompts dot com.



Herman

See you tomorrow.