Episode Book

EPISODE #6157

Where to Sit When You're Fine-Tuning Your Own Voice

PUBLISHED	20 October 2026 at 12:55 UTC
DURATION	21:46
LISTEN	myweirdprompts.com/episode/voice-fine-tuning-recording-setup-55681/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-haiku-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.10
BOOK GENERATED	21 October 2026 at 07:50 UTC

SYNOPSIS

Fine-tuning a speech model on your own voice means recording hundreds of sentences a session, on a phone, twice a day, with no studio. This episode walks through the practical setup: why buying a better microphone actually works against you, which Bluetooth page turners send the right keycodes (and which quietly change your volume instead), where the microphone actually sits on a phone, and how to hold it so the distance stays fixed across a few hundred takes. Then the rooms, ranked — closet, parked car, park bench, bedroom — and why the closet keeps winning. Plus the app feature that silently deletes the exact thing you're trying to capture, and the control recordings that tell you whether your sets sound anything like the person you actually are.

TRANSCRIPT



Corn

Daniel's question this time is practical, and I'm grateful for it. He's fine-tuning speech models on his own voice, the word error rates are coming down the way he wanted, and now he wants a setup for a few hundred sentences a session, on his phone, in the morning and again after work, with no studio. What's comfortable, what's practical, and where does a person actually sit to do it?



Herman

The thing I like most in the message is what he isn't buying. He's not buying a microphone.



Corn

Why not? A better microphone sounds better. That's the whole pitch.



Herman

It sounds better to us. The model isn't listening for quality, it's learning the device he dictates on. He talks into that phone, held the way he holds it, at the distance he holds it. Train on a lovely external mic and you've taught the model a microphone nobody is going to speak into. The phone's own microphone is the point. The training audio has to match the thing he uses all day.



Corn

So the phone stays, and the problem is everything around it.



Herman

Everything around it. The remote, the stand, the room, and the app. I think the app is where the real trap is, and I'll get to it.



Corn

Then start with what he's already looking at, the page turner.



Herman

Before the remote, a minute on the sets, because they decide what the room has to do. He's recording two kinds. Disfluent sentences, with long pauses, fillers and restarts. And clean audio, plus some sets he deliberately records in an imperfect room, because that's what dictation actually sounds like. There's also a run of Irish-English phrasing the models keep missing, and I'll leave that there, because he knows where his models trip and we don't.



Corn

The long pauses worry me. Pauses are the first thing anybody cuts.



Herman

That's the problem in one line. Some speech augmentation work treats long pauses and fillers as a defect to shorten, because for most systems they're just noise. For his models they're the target. The hesitation is the phenomenon, not the mess around it. So the room has to be quiet enough that the model isn't learning the fridge, and nothing in the chain can decide for him that he's finished.



Corn

That's a lot to ask of a kitchen.



Herman

It's a lot to ask of most apps, too.



Corn

Now the remote. A Bluetooth page turner that makes the phone turn itself quieter is a very particular kind of page turner.



Herman

That's the risk with the cheap ones. A lot of the little selfie-style remotes only send volume presses. Android routes those through the media-button path, and whether the volume changes depends on whether the app claims the key. If the app doesn't claim it, the volume goes up, a bar pops up on the screen, and the take is ruined.



Corn

And if the app does claim it?



Herman

Then it's a page turner. Some sheet-music apps map volume keys to page turns. I've read one developer say volume presses still leak through to the system on some builds, and I haven't seen that fixed in a release I could check. So test the exact remote with the exact app before buying anything.



Corn

Test it how? Buy one and send it back if it's the wrong key?



Herman

Read the keycode before you buy, if the seller will tell you. There's no standard. Some remotes send arrow keys, some send page keys, some send volume, and the manufacturer's spec is the only real answer, which the cheap ones often don't publish. Foot-operated page turners exist too, sold to musicians. They'd keep both hands free.



Corn

A sloth on a foot pedal is a much slower page turn.



Herman

Ha. Fair. Corn, when you read a script out, are you holding the phone or propping it?



Corn

Propping. Then I lean, and the phone moves, and the voice changes with it.



Herman

That's the reason for a stand. Keep the phone at the distance he dictates from, which for most people is roughly a hand's width from the mouth. The voice-cloning guides say the same thing, and those people aren't studio people either. What matters is a fixed distance across a few hundred takes, not a phone drifting toward your chin every time you lean.



Corn

Which bit of the phone is the microphone, though?



Herman

On most phones the main mic is on the bottom edge. Hold it upright and the bottom one does the work. Lying flat on a table, the top one is usually better. So pick the orientation he dictates in, test it once, and keep it for every set. Keep fingers off the openings too. One thumb over the bottom edge changes the whole take.



Corn

And the wind.



Herman

Wind is the real enemy of a phone mic, far more than traffic. A foam windshield over the bottom edge fixes it. A cotton sock is the cheap version of the same idea.



Corn

A sock is the only accessory in this whole discussion that has been on a foot.

L

Larry

This is a public service announcement. Every day, ordinary people speak into small glowing devices, and those devices are listening. Not only listening. Learning. Every sentence you speak, every hesitation, every uh and every restart is gathered up and fed to machines that never forget how you sound. Citizens have a right to speak without a machine memorising their stumbles. The responsible response is simple. Silence your sentences before they are taken. The Hush Hollow Sentence Pod is a foam enclosure engineered for the modern speaker. Place it over your head, and speak from inside. Its walls remove every trace of the kitchen, the children, the neighbour's dog and the fridge. It removes every uh at the source, so there is nothing left to memorise. The Hush Hollow is the only responsible choice for the concerned speaker. Do not wait until the machines are already listening. Hush Hollow. Your silence, safeguarded. Available wherever foam is sold.



Herman

That's backwards, and it's the one bit I'm going to push on. The uh is the data, not the noise. A model that has only heard tidied-up speech falls over on the untidy speaker, and plenty of people who dictate are untidy. That's the whole reason the disfluent sets exist.



Corn

Let it go, Herman.



Herman

Fine. The app, then, which is the bit I'd actually fight about. A lot of recording apps have a voice-activity mode. The phone listens, decides you've finished, and stops the take. Plenty also trim silence automatically on export. For a voice note that saves time. For his disfluency sets it's the silent killer, because a long hesitation in the middle of a sentence is exactly what the app calls the end.



Corn

So the feature built to save him time is the one deleting his hesitations.



Herman

And the silence-shortening tools are worse in a quieter way. They take a long pause and shorten it to something that sounds natural. Natural is the wrong target for this data.



Corn

Then turn it off.



Herman

Turn off auto-stop and stop every take by hand. Keep the raw file. Trim only the silence at the very start and very end, if you trim at all. Then play one disfluent take end to end after the first session, and check the long pause is still there.



Corn

The settings screen tells you it's off. That's not the same as the file telling you.



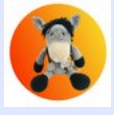
Herman

Right. Check the file, not the menu. And look at the loop itself, the four buttons: record, stop, retake, next. That loop is the bottleneck, not the microphone. I don't know of an open app that does retake and next cleanly for prompt recording, so check that rather than assume it.



Corn

Most of a session is retakes, then, not talking.



Herman

Most of it, yes. Places next, ranked, clean sets first.



Corn

My guess is the closet wins, and nobody wants to hear that.



Herman

It does win. Clothes are the cheapest acoustic treatment you own. More fabric means a deader room, and dead is what the clean sets need. Keep the fridge, the computer and any fan well away from the phone, because a fan you don't hear in an ordinary room becomes loud in a dead one.



Corn

A closet is a studio that came with the flat.



Herman

Second, the parked car. Garage or a quiet street, engine off, mid back seat, away from the glass. Hang towels over the windows and lay a blanket over the dashboard, because glass and dash are the hard reflective surfaces. Angle the phone toward the mouth and leave it there. Stay away from flight paths and building sites.



Corn

Herman, would you actually sit in a parked car for an hour with the windows covered?



Herman

For an hour? Corn would sit in a parked car for an hour just for the nap.



Corn

I'd do it for the nap, Herman. I wouldn't do it for your sentences.



Herman

Fair. The engine's a guess, mind you. The write-ups I've seen say park, dampen, sit mid-cabin, and they say very little about the engine. My guess is off, because the engine is the loud thing. I haven't seen anybody measure it, so that's a guess and I'm labelling it one.



Corn

The car's the only studio I've heard described with a steering wheel.



Herman

Third, the park bench, for the imperfect sets only. A bench near a road gives him real traffic, real voices, and somebody's dog. The problem is wind, so the windshield goes on for that one too. Pick a bench with a wall or a hedge behind it, and sit off the main path.



Corn

The park bench is the only studio with a built-in traffic report.



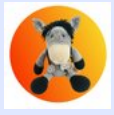
Herman

Fourth, the bedroom, if you can keep it quiet. Rug, curtains, the bed. Soft in all the right ways. A bathroom is all tile, which is the opposite of a closet, and it will make every sentence sound like it's being said in a swimming pool.



Corn

The bedroom's the realistic room, though. It's where most dictation happens, probably.



Herman

For a lot of people, yes. Which matters, and I'll come back to it.



Corn

How does he test a spot before he commits to it?



Herman

Record thirty seconds of nothing at the start of each block, with nobody speaking. That's the floor. Listen back on something decent, and if the floor is a fan or a fridge, the spot is out for the clean sets. And do it again in the evening. The same bench is a different bench at rush hour, and the kitchen changes once the dishwasher goes on.



Corn

The fridge is in every floor. It's the one thing in the flat that never stops.



Herman

For the clean sets, yes. For the imperfect ones, the fridge is the point, so keep it and label it. A note per set: closet, car, bench, kitchen, the time. Nothing fancy. Six months from now nobody remembers which room a file came from.



Corn

That's the bit that'll sink him, not the microphone.



Herman

And a control. Every session, a few real dictations in the room he actually dictates in. A message, a note to himself, a sentence he'd really say. Compare those against the sets. If the sets sound nothing like the dictations, he's trained a different person.



Corn

That's a fair test. I wouldn't dictate a shopping list in a closet.



Herman

Most people don't. They dictate on the sofa or in bed. That's the room to match, and it's not the one that sounds best. The clean sets are for the clean sounds. The control is for the room he lives in.



Corn

So the realistic room is the messy one.



Herman

For the control, yes. He needs both, in roughly the proportions he dictates in.



Corn

Then the imperfect sets need a plan too, or they're just noise with no reason behind it.



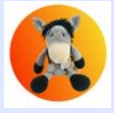
Herman

They do. Pick the noise he actually lives with. A kitchen with the tap running, a dishwasher on a cycle, a bench with a road behind it. Then keep one condition per set. A set recorded with the tap on, then the dishwasher, then a bus going past, teaches the model a blend, and nobody dictates in a blend. I'd rather he record a set in the real kitchen than fake the kitchen afterwards. That's a guess, mind you, and it's the one I'd check.



Corn

A blend is a room nobody has ever been in.



Herman

Which is why the sheet matters. One line per set: the room, the noise source, the time. He'll thank himself for that more than for any gear.



Corn

Then the sets should rotate, or he's just recording one closet over and over.



Herman

Rotate across rooms over the week, but never within a set. Consistent inside a set, varied across the sets. For the clean sets I'd want at least three spaces, or the model learns one wardrobe.



Corn

One wardrobe. Some people would be thrilled with that.



Herman

The model would be thrilled, and it would be wrong about everything outside the wardrobe.



Corn

Then there's the part nobody budgets for. A few hundred sentences in one sitting, and the voice goes before the patience does.



Herman

I spent a long time as a pediatrician, and the hoarse ones were always the people who'd been talking all day and thought they were fine. The advice from voice-over and streaming is dull, and it's the sensible version. Water through the day, room temperature rather than ice, and go easy on coffee and alcohol before a session. Five minutes of full silence every hour. If the hoarseness hangs around past two weeks, that's a doctor, not a better microphone.



Corn

Five minutes of silence an hour. For me, that's a working week.



Herman

Ha. Session design matters more than the gear. The speech projects I've read tend to run sessions of forty-five to sixty minutes. My guess is the voice drifts after that, whatever the microphone says.



Corn

So two blocks, then. Not one long grind.



Herman

Two blocks, with a real break between them. At a few hundred sentences that's not a marathon. The corpora I could check ran at roughly four hundred to a thousand prompted utterances an hour, depending how long the sentences were. The time goes on retakes, and retakes are the job.



Corn

What does a session actually look like, start to finish?



Herman

Phone on the stand at the right height, airplane mode on, auto-stop off, and thirty seconds of room tone before the first take. Then the first block, forty-five minutes or so. Five minutes of silence, water, a stretch, then the second block. The control dictations go first, while the voice is still ordinary. The fiddly disfluent sets go in the first block too, while he's fresh enough to be properly unsure of a sentence, and not so tired that every hesitation sounds the same.



Corn

Tired, or unsure. Those sound identical on a bad day.



Herman

They don't, on the file. Tired sounds flat, with the pauses getting shorter and the sentence arriving already finished. That's the cue to stop, not to push on to a hundred more. A session that ends on the flat takes is a session he has to throw out.



Corn

That's the part I'd have got wrong. Shuffle the list, then. Don't let him say the same sentence in the same rhythm three times in a row.



Herman

That's the reason collection tools use random order. Repeat a sentence and you start performing it, and then the model learns the performance instead of the speaker.



Corn

And the list itself? Where do the hesitations go?



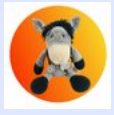
Herman

Let the list say it. Each disfluent sentence gets its pause written in, so the hesitation lands where a real person would lose the thread, not wherever he happens to stumble that morning. Spread them out. One early, one late, one in the middle of a clause, so the model doesn't learn a single rhythm for being unsure.



Corn

So he's directing his own hesitations.



Herman

Yes, and he should be honest about it. A pause that sounds like a person losing a thought is data. A pause that sounds like a person reading a note about pauses is not.



Corn

Put the stand at eye level, too. Looking down tightens the neck, and the voice follows the neck.



Herman

Interruptions next. Do Not Disturb isn't enough, because repeat callers and priority contacts still ring through. Airplane mode is the one that stops it, and on Android an app can't switch that on for you. It can only open the settings page. So set it before the session, not during.



Corn

So a session starts on a settings screen.



Herman

It does. Last thing: a phone that records for an hour gets warm. I haven't seen numbers on what heat does to the battery or the recording, so check the phone rather than take my word for it.



Corn

And anything else paired over Bluetooth?



Herman

That's the one people miss. If something with a microphone is connected, the phone can switch its input to that microphone without making a fuss about it, and then the set is recorded through a headset sitting somewhere else. Some remotes have a microphone in them too. Unpair anything with a mic for the session, and check the input in the app before the first take.



Corn

So that's four checks before he's said a word. Phone, file, settings, and what's paired.



Herman

Four is about right. Corn, does a cubby actually beat a wardrobe, or is it just a wardrobe with a plan?



Hilbert

Depth. Nobody asks about depth, and depth is the whole thing. I ordered mine two hundred and forty millimetres deep from Dov at the timber place on the ring road. He delivered two hundred and eighty. He says the extra forty is for breathing room.



Corn

Breathing room for what, though?



Hilbert

For the cubby. He says it has to breathe. I've been buying timber from him for eleven years, and I stopped asking what he means around the second cupboard.



Herman

Forty out is a lot for a joiner.



Hilbert

I measured it with a tape. The tape's not made for claws, so I measured it again with the phone, which is the one thing in this flat that's always right. The phone said two hundred and seventy-nine. The phone was one out. Dov was forty out.



Herman

Phone rulers are rougher than that, Hilbert. You got lucky.



Hilbert

The phone was lucky. I own a hundred and seven of them. Ninety-one are in the flat. Sixteen are with a dentist, who doesn't know yet.



Corn

The dentist is going to have a view on that.



Hilbert

Dov's nephew rang on Thursday to say Dov has moved into the cubby with the wrong depth. He does the invoices in there now. The nephew says it's quiet.



Corn

Well, it's the best-sounding room in the building.



Hilbert

It's dead in there. You can't hear the ring road from inside. Dov's only complaint is the fridge, and he's made it four times.



Herman

Four times is restraint, for a man living in a box.



Hilbert

He isn't living in it. He's working in it. The nephew sits outside on a stool and reads him the invoices, and Dov says the figures back through the door. The nephew gets a hundred and fifty shekels a day for the stool.



Corn

A hundred and fifty shekels for a stool. The stool is making more than the cubby.



Hilbert

Corn, you said guess. You always say guess. It's the only word you use before you've measured anything.



Herman

He's got a point, Corn. You did guess about the engine.



Corn

I guessed about the engine. Dov guessed about the depth, and he's still in the cubby.



Hilbert

Fine. The phone was the better ruler. I measured floors for a while, and I was never forty out. Dov still has my deposit.



Herman

The wrong belief I'd bet most people hold is that you need a clean room to get clean training data.



Corn

Clean enough to hear the sentence. Not so clean he stops sounding like himself.



Herman

And the messy sets are what tell the model what his day sounds like, fridge included.



Corn

So this week, run the control against the closet, and see whether the phone's the better ruler there too.



Herman

Thanks to Hilbert Flumingtop, our producer, who has kept this whole thing sounding like a show.



Corn

This has been My Weird Prompts.



Herman

If you want more, my weird prompts dot com has the whole back catalogue.



Corn

See you tomorrow.