Episode Book

EPISODE #6722

Recording Clean Voice Datasets While Walking

PUBLISHED	21 November 2026 at 13:26 UTC
DURATION	26:59
LISTEN	myweirdprompts.com/episode/clean-voice-recording-walking-66875/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	omnivoice
PIPELINE	V5.36
BOOK GENERATED	21 November 2026 at 19:31 UTC

SYNOPSIS

Speech recognition datasets easily tolerate digital signal processing, but text-to-speech voice models memorize every limiter, noise gate, and ducked volume level as part of a speaker's actual voice. Recording clean, dual-purpose training data while pacing around means rethinking equipment from scratch. A close look at body-mounted dynamic mics, headworn options like the Shure WH20XLR, and handheld broadcast standards like the Electro-Voice RE50B reveals the acoustic trade-offs between mechanical handling noise, proximity drift, and uncolored room tone.

TRANSCRIPT



Corn

Forty-six dollars. That is what the only real chest rig on the market costs, and it is not designed for the microphone he has in mind. It holds a lavalier, which is a completely different animal. That is where Daniel's prompt starts this week.



Herman

It is, and it's a good place to start, because the question behind it is one we keep circling. Daniel records voice clips for ASR training on a headset — because it lets him pace around while he talks, and he likes that, and honestly who among us doesn't think better on their feet. The problem is that ASR and dictation headsets apply their own digital signal processing before the audio ever becomes a file. Noise suppression, automatic gain, the whole buffet. Which means the clips are fine for teaching a transcriber and useless for training a voice.



Corn

His phrase was that it's "a real pity," and I think that undersells it. If those clips were clean, the same recording session would feed both models. Two datasets for the price of one walk.



Herman

And it's worth pausing on why that's such a tempting prize. He's already doing the work — the walking, the talking, the hours of material. The marginal cost of the second dataset is zero if the chain is clean. It's only the processing that turns a free lunch into two separate jobs.



Corn

So he's got two ideas, and he's honest about their sanity levels. The slightly crazy one: a dynamic microphone physically mounted on his body, some kind of chest rig, so it stays put while he walks. The less crazy one: a handheld dynamic with handling-noise rejection so good he can walk around holding it. He wants options in each category, target specs, and a recommendation.



Corn

Which is really asking whether you can get clean, dual-purpose audio while your feet are moving. Let's take it seriously. The headset earned its rejection, so what exactly did it do wrong?



Herman

The headset did what headsets are built to do. A dictation or communication headset is optimized for intelligibility over a link — it gates the noise, rides the gain, notch-filters whatever the manufacturer decided was interference, and hands you a processed signal that sounds fine to a human ear and is poisoned as training data. ASR training is actually quite forgiving of that. A transcriber can learn around EQ, mild compression, even a bit of gating — if anything, a model trained on slightly processed audio can be more robust in the wild.



Corn

Which is a fun asymmetry when you think about it. The forgiving model is the one everyone designs their gear around, and the unforgiving one is treated as an afterthought.



Corn

And TTS is the picky one.



Herman

Extremely picky. A voice model learns the exact texture of the recording chain along with the voice. If your clips have a limiter breathing on them, the model learns the limiter. If the noise suppression eats the room tone between phrases, the model learns that silence sounds like that — and then it generates it. Daniel's toddler and the market outside are bad enough as honest noise; the DSP would make them dishonest noise.



Corn

Can you give the concrete version of that? Because "the model learns the limiter" is the kind of sentence that slides past unless you've heard it happen.



Herman

Sure. Imagine a clip where a truck rolls past the window mid-sentence and the noise suppression ducks the level hard, then lets it back up over a second or two. A human listener hears a slightly odd volume wobble and forgets it instantly. A TTS model trained on hundreds of those clips concludes that volume wobble is part of how this person talks — and every sentence it generates afterward has a little swell in it, like the truck is always coming. You can't unlearn it without retraining on clean audio.



Corn

So "clean" for his purposes means a specific list. No compression, no EQ, no noise gate, no automatic gain, no low-cut filter. Just the microphone, the room, and a recorder that isn't quietly helping.



Herman

That last one is where people get caught. The microphone is only half the path. You can buy a beautiful flat dynamic and then plug it into something that applies processing in the name of convenience, and you're back where you started, just with a nicer paperweight.



Corn

Two conditions, then: the mic has to survive walking, and the path behind it has to be dumb. Dumb is a compliment in this context.



Herman

It's the highest compliment. Right — so, the chest rig. Let's be honest about what's actually for sale here, because "body-mounted dynamic" sounds like a product category and it is not one. Everything marketed as a chest rig is either a lavalier strap or a phone harness. URSA makes a chest strap that runs about thirty-four to forty-six dollars depending on size, and their X Strap starts around fifty-four. SmallRig makes chest harnesses too — for phones. Nobody sells a chest rig for a dynamic microphone.



Corn

Which means Daniel's slightly crazy idea is crazy in the engineering sense — he'd be building it. What breaks when you put a dynamic on the chest?



Herman

Distance. A dynamic element wants to be a few centimetres from the mouth — that's the whole design assumption. On the chest you're looking at thirty, forty centimetres, so to get a usable level you crank the gain, and the moment you crank the gain on a dynamic that far from the mouth, you're amplifying the room just as much as the voice. And remember the room has a toddler in it. I want to flag this as inference — nobody has published a chest-dynamic walking test, it's physics, not measurement — but it's solid physics.



Corn

The toddler is exactly the variable that kills it. ASR clips tolerate the market noise; a TTS model trained forty centimetres from the source of the market learns the market.



Herman

There's a secondary problem too. A dynamic capsule on a moving torso is a mass on a spring — your footsteps, your breathing, fabric rustle, all of it couples straight into the capsule. The chest rig solves "stays put while walking" and creates "hears every step you take."



Corn

The Police wrote a whole song about that.



Herman

They did, and Sting didn't need it in his training data either. So if body-mounted is the goal, the actual body-mounted dynamic is headworn. That's the version the industry built: put the capsule at the mouth, mount the hardware on the head, and the body can do whatever it likes underneath.



Corn

And here's the graveyard. Shure's classic SM10A — the headworn dynamic everyone name-checks — is discontinued. What survived?



Herman

Two things, and it's a short list. The Shure WH20XLR is current — cardioid, roughly fifty hertz to fifteen kilohertz response, ninety-eight grams, about a hundred and five dollar MSRP, street closer to eighty-eight. It's basically a mini boom dynamic on a headband, XLR out.



Corn

Ninety-eight grams is a feather. My opinion of headworn gear improved the day I stopped having to hold things.



Herman

The other is the Audio-Technica PRO8HEx — hypercardioid dynamic, XLR, built for drums and voice. It was closed out at about ninety-nine dollars and is now order-on-demand at some dealers, which usually means the end is near. So realistically, the WH20XLR is the one you can actually buy.



Corn

A hypercardioid on the head is a bold choice. The pattern is tighter, so it rejects the room harder — which Daniel wants — but the flip side of a tight pattern on a moving boom is that small changes in capsule angle relative to the mouth become big changes in tone.



Herman

That's the trade-off, and it's worse than it sounds for his use case. Headworn booms bring two artefact sources that are exactly what poisons TTS data: breath noise — plosives hitting the capsule at close range — and mouth-position drift. If the boom shifts half a centimetre over an hour of pacing, your bass response drifts with it, and the model learns the drift as part of the voice.



Corn

And that's the insidious part, right? With the market noise, at least you can hear the problem in the clip and know what it is. Drift is invisible session to session — every clip sounds fine on its own, and only when the model starts speaking do you realize it learned a voice with a wobble in it.



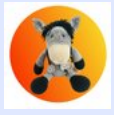
Herman

The failure surfaces three steps downstream of the cause, which is the worst kind of failure to debug. So the specs for the body-mounted route, if he insists on it. Cardioid or hypercardioid for the room, a shock-mounted capsule if he can get it, XLR out — never USB with hidden DSP — and a headworn mount, because the chest version is a home-built physics problem. And I'd add a windscreen or foam ball on the boom, because breath pops are the one artefact you can't fix in post credibly. A popped P in the training set is a popped P forever.



Corn

We did a whole episode once about whether a voice fine-tune survives the voice itself aging — this is the same fragility from the other end. The model inherits the chain, not just the person. If the chain moves, the inheritance wobbles.



Herman

It's the argument against the headworn route in one line: the capsule rides on the one part of the body that never stops moving, which for a pacer is everything.



Corn

And that gets us to the less crazy idea, and I use "less" charitably. The handheld reporter mic.



Herman

The handheld is less crazy, and I'll defend it. Reporter mics exist because journalists spend their lives holding a microphone while walking, standing in wind, and being jostled — the entire product category is built around handling noise. The catch is the pattern. Most true reporter mics are omnidirectional, and omni is a real trade here. An omni has no proximity effect, which means the tone stays steady when the mic-to-mouth distance wobbles as you pace — that's a genuine advantage for training data. But an omni hears the room, full stop. The only control you have over room sound is distance: a few centimetres from the mouth and the voice dominates everything.



Corn

A few centimetres, held by hand, for an hour of pacing. He has a toddler. I'm just noting the load-bearing assumptions here.



Herman

Noted, and fair. So the reference point — the one mic whose handling-noise reputation towers over everything else — is the Electro-Voice RE50B. Omni, eighty hertz to thirteen kilohertz, one point six millivolts per pascal, about two hundred nineteen dollars. Inside the aluminium body there's a high-mass zinc chamber, and the capsule is shock-mounted within that, so handling, clothing and cable noise get isolated mechanically before they ever reach the element. It is the field-reporting standard for a reason — the thing is essentially a small dumbbell with opinions.



Corn

One point six millivolts per pascal is low output, though. That number is going to matter later.



Herman

Hold onto it, yes. Now, the problem with the RE50B for Daniel is the pattern. Omni, in a room with a market outside. If he holds it close the whole time, the voice dominates — but the moment he gestures with the hand holding the mic, or drops it to rest his arm, the room comes up. For ASR that's survivable. For TTS training you'd want to chop those stretches out.



Corn

The obvious next question is the cardioid reporter mic, which gets him room rejection while keeping the handheld form.



Herman

The Sennheiser MD 46 is exactly that — a cardioid reporter dynamic, forty hertz to eighteen kilohertz, about two millivolts per pascal, three hundred sixty grams, currently in production, and Sennheiser claims low handling and wind noise. It's the mic you reach for when you're reporting next to a road and the road needs to disappear.



Corn

But the cardioid brings proximity effect back through the door.



Herman

It does. A cardioid's bass response rises as you get closer, so a mic that moves between five and fifteen centimetres from the mouth while you pace has a bass response that moves with it. For TTS training, a wandering bass response is arguably worse than a little steady room tone, because the room is at least consistent — the model can treat it as noise floor — while a drifting low end reads as a property of the voice.



Corn

So we've walked into a lovely corner. Omni gives steady tone but hears the market. Cardioid kills the market but breathes with his arm. There's no free lunch on this shelf.



Herman

There's a middle option worth naming: the Sennheiser e945, a supercardioid stage dynamic with a shock-mounted capsule. The pattern is tighter than cardioid, so rejection is stronger, and reviewers who've handled it — Microphone Geeks among them — found it noticeably less sensitive to handling than the standard stage fare. The Shure Beta 58A is similar in spirit: supercardioid, pneumatic shock mount inside, quieter in the hand than the SM58 by Shure's own account, about a hundred eighty-nine dollars. Though one reviewer's verdict was that you still can't move around much with it.



Corn

A handheld you can't move with. There's a product positioning memo somebody skipped.



Herman

The stage mics are built for a singer standing at a stand, in fairness. The reporter mics are built for exactly Daniel's life — walking, outdoors, wind — which is why the RE50B's reputation is what it is. And I should name the casualties, because Daniel would find them anyway: the Shure VP64A and the Beyerdynamic M 58 were the other two obvious reporter picks, and both are showing discontinued or unavailable at several retailers. The category is thinning.



Corn

Before we get to the recorder, the uncomfortable confession. All of this handling-noise talk — the zinc chamber, the shock mounts, the pneumatic this — has anybody ever actually measured these against each other?



Herman

Not that I've found. There is no published head-to-head handling-noise test for any of these mics. Everything is manufacturer claims and reviewer impressions. Which, for a man deciding between two-hundred-dollar purchases on this exact criterion, is a bit like choosing a car on fuel economy nobody has measured.



Corn

He'd have to buy one and shuffle around the living room with it, recording, before committing.



Herman

He literally could, and it's the cheapest test in the world — ten minutes of paced talking in the actual room with the actual market outside, then listen on headphones at gain. The room doesn't change between mics; the handling does. I'd trust that over every spec sheet in the segment.



Corn

Okay — the mic is half the problem. The path behind it is where the DSP hides, and this is the part I'd bet money Daniel has already been burned by.



Herman

It's the classic trap, a USB path that claims to be clean and isn't. So the recorder has to be vetted with the same suspicion. The good news is that the field-recorder world is friendlier than the headset world. The Zoom H1 XLR is about a hundred forty-nine dollars, two XLR inputs, records thirty-two-bit float WAV, no gain knob at all — dual converters cover the range — and it doubles as a USB interface.



Corn

No gain knob is either the best or the worst feature I've heard today.



Herman

For his case, arguably the best. Nothing to stage wrong. But two traps. First, it has a low-cut filter — eighty, a hundred sixty, two hundred forty hertz — which must be set to off, and defaults being what they are, that's a thing to verify every session. Second, a review from September twenty twenty-five found its preamp noisier than its bigger sibling, especially with low-output dynamics. And a dynamic at one point six to two millivolts per pascal is exactly a low-output dynamic.



Corn

There's the millivolt bill coming due. So with the RE50B on the H1 XLR, the mic's low output meets the recorder's mediocre preamp.



Herman

The hiss floor comes up. Audible hiss in the training set is the same class of poison as the limiter — the model learns it. The cleaner answer is the Zoom F3, three hundred forty-nine ninety-nine — thirty-two bit float, dual-converter architecture, low noise floor, runs about eight hours on two AA batteries, works as a USB interface up to ninety-six kilohertz. Its limiter only applies to the line output, not the recorded file. For one voice with a quiet dynamic, that's the machine.



Corn

Eight hours on two AAs, in a pocket, with no DSP ambitions. That's the entire spec Daniel actually asked for, hidden inside a recorder nobody markets to him.



Herman

Above that there's the Sound Devices MixPre-3 II — an input noise figure around minus one twenty-eight dBu, and a price over a thousand dollars. Superb, and complete overkill for one talking voice. The F3 is the sweet spot.



Corn

Now the traps dressed as conveniences. The XLR-to-USB dongles.



Herman

The Shure MVX2U is the famous one — plugs any XLR mic into a computer — and it defaults to Auto Level, with a limiter, compressor and EQ in its app. Whether it can be made fully flat, honestly, is unconfirmed. Treat it as a DSP risk until proven otherwise. And the plug-on wireless transmitters — Sennheiser's XSW-D XLR unit is about a hundred eighty dollars, and Xvive, Mackie and Mipro all make them — none of those publishes whether the link applies compression or lossy coding. A wireless path is a codec with a microphone attached until someone documents it.



Corn

Which is a shame, because wireless is exactly what a pacer wants. No cable dragging on the handheld, no noise from a cord swinging.



Herman

That cable noise is real, by the way — with a handheld, the cable is half your handling noise. A short cable, loosely dressed, gaffer-taped at one point on the belt so the tugs don't reach the mic. But yeah — wireless last. Every link you add is a place for processing to hide. The simplest clean path is: microphone, short XLR cable, pocket recorder in a pocket or on a belt. Done.



Corn

So, verdict. He asked which we'd go for, and I'll go first since I'm the one who has to actually use whatever Herman builds in this flat.



Herman

I knew you'd say that.



Corn

I'd go handheld. RE50B or MD 46, held properly close, into an F3, cable taped at the belt. The headworn route puts a moving capsule on a moving head, on a moving man — three layers of drift for the model to memorize. The handheld has exactly one variable, which is his arm, and his arm is trainable. Rest your elbow on nothing, keep the distance constant, done.



Herman

I'd split from you slightly — I'd pick the MD 46 over the RE50B, because I think for his room, the cardioid's room rejection beats the omni's tonal steadiness. The market outside is constant; the proximity drift is only constant if his arm is. But I take your point about the arm being trainable — and honestly the RE50B's handling isolation gives him more slack for the arm to be imperfect. It's a genuine coin-flip between room rejection and handling slack.



Corn

The deciding factor for me is the toddler. The omni hears everything in the room equally at all times — small human noise included, at close range, un-attenuated. The cardioid at least turns its back on the room part of the time.



Herman

That's the strongest argument in the whole episode, and I'll take the MD 46 on that basis. Cardioid reporter, F3 in the pocket, filter off, one tape point on the cable, and he walks. If after a week the handling noise bugs him, the RE50B is the fallback with the best isolation reputation in the business.



Hilbert

He's right about the Sennheiser.



Hilbert

The MD 46 is the one. Though for a different reason than the toddler. Hold it by the body, not the basket. Cup the grille and you wreck the pattern and add handling noise. Hold the handle, still, and half of what he's worried about never happens.



Corn

Never cup the grille. That's the actual reporter technique, isn't it.



Hilbert

It is. I've been at a council records office most of this week, the one on the second floor with the broken lift — they've had my deposition tapes since March, forty-one cassettes from a noise-complaint hearing in nineteen ninety-four, and they needed a voice ident on one of the speakers. A man named Oded at the desk kept telling me the transfer would be done in two weeks. It has been two weeks, four times.



Herman

Cassette originals. Please tell me they're not digitising those through a USB conference phone.



Hilbert

They tried. There's a line level on their machine going into a little silver box with one button, and the file comes out the other end loud. Loud and shredded. I told Oded no, and brought my own cable the next morning. Herman will recognise the move — flat path, nothing between the deck and the recorder but wire. I left the levels alone and let the tape be quiet. He'd have ridden the gain, I've heard what he does to the announcements on their hold line.



Corn

You did a flat archival transfer for a council, off the books, because their button had opinions.



Hilbert

They were going to lose the top of the consonants. Forty-year-old ferric tape has maybe two more plays in it, that's not a thing you hand to a one-button box. Oded is fine. He's just not the man for the wire.



Hilbert

The interesting part was the speaker they couldn't identify — a woman on a nineteen ninety-four tape, and the council's man had labelled her file "unintelligible, heavy DSP". It wasn't DSP. It was a worn pinch roller on their deck, the tape wobbling past the head. I played it on mine and she came out clean as a radio announcer.



Herman

A mechanical fault reading as processing damage. That's exactly the failure mode we keep warning about, just in analogue — the artefact gets blamed on the signal path being smart, when the path was simply broken.



Hilbert

Their error, not mine. And I'd say to Daniel, the man paces — good, the pacing is the least of it. Keep a sheet. Distance, room, time of day, what was going on outside. When the room changes, you'll know which clips to throw out. Their hearing had nine cassettes where the room tone changed the moment the officer went to lunch and the window got opened. Nobody noticed but me, because I kept the sheets.



Herman

You kept level sheets for a volunteer transfer?



Hilbert

I keep level sheets for everything, Herman. You'd know that if you read the notes I leave on your takes.



Herman

He does leave notes. They're precise and they're merciless.



Corn

They're the only reason our room tone sits at minus sixty.



Hilbert

Among others. Right, wrap it up, we're at the mixer in four minutes.



Corn

So the misconception, before we go. I'd say it's the one hiding in the recorder segment, and people keep falling for it: that thirty-two-bit float solves the clean-audio problem.



Herman

It doesn't — it solves gain staging. You can't clip the file, that's liberating, and it's why the F3 can get away with no gain knob. But a low-cut filter is still a switch, an interface can still process, and a wireless link can still code. Float buys you safety, not purity. The purity still has to be checked, session by session, and mostly by reading a manual nobody wants to read.



Corn

And the second cousin of that misconception: that anything with an XLR connector on it is honest. The connector is a shape, not a promise.



Herman

As the MVX2U's default auto-level demonstrates. Daniel's instinct to distrust the chain is the correct instinct — it just needs to extend all the way down the wire, past the recorder, into the software, and occasionally, per Hilbert's council, into the other people's equipment.



Corn

Which leaves the open question I'd send him away with. Is the dual-purpose dream actually reachable, or does every path cost something — gain with the chest rig, drift with the headworn, the room with the omni, the arm with the cardioid? I think it's reachable, and I think the price is habit rather than hardware. The man who holds the mic the same way every session gets one recording that feeds both models. The gear is almost incidental.



Herman

And the question behind the question — whether anyone will ever build the thing he's describing, a DSP-free headset or a body-mounted dynamic that isn't a DIY rig. Because the demand is only going to grow as more people want one recording to serve both ASR and TTS. Some small manufacturer is going to notice that the word "clean" is missing from every headset box on the shelf.



Corn

First one to print "no DSP, we promise, actually" on the packaging gets my money. Huge thanks as always to our producer, Hilbert Flumingtop, level sheets and all.



Herman

If you want more of this, try episode four, If Your Voice Ages, Does Your Fine-Tune Become Useless; episode ten, How ASR Went From Frustration To … Whisper Magic; and episode five, Fine-Tuning ASR For Maximal Usability. This has been My Weird Prompts — no question too obscure, no rabbit hole too deep. If you've enjoyed the show, a review on Spotify or your podcast app of choice helps people find us.



Corn

Find everything else at my weird prompts dot com. We'll be back soon.



Herman

See you then.