Episode Book

EPISODE #6636

The Five-Step EQ and Compression Chain for Any Voice

PUBLISHED	16 November 2026 at 21:19 UTC
DURATION	23:44
LISTEN	myweirdprompts.com/episode/voice-eq-compression-tts-mastering-31368/
HOSTS	Corn, Herman, Dorothy, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	omnivoice
PIPELINE	V5.34
SOURCES	25 cited by the research (listed at the end)
BOOK GENERATED	16 November 2026 at 22:22 UTC

SYNOPSIS

Mastering human and synthetic voices relies on the exact same five-step toolkit: an 80 Hz high-pass filter, a 200–400 Hz mud cut, a 3–5 kHz presence lift, a 5–9 kHz de-esser, and gentle compression. The order matters—EQ must come before compression so desk rumble doesn't trigger the compressor. Beyond the numbers, this conversation unpacks why Shure SM7B bass-boosting fatigues car listeners compared to Neumann U87ai clarity, and why TTS engineering feels newer: it fixes model artifacts rather than physical room capture.

TRANSCRIPT



Corn

We've spent a lot of recent episodes down in the TTS weeds, EQ passes and adapters and so on, and Daniel wants to zoom out. His question: what do we actually mean by working with voice material, human or synthetic, to make it listenable? He notes that voice engineering for TTS is in its infancy, while EQing human voice talent is decades old, and he wants to know why that gap exists.



Herman

And he's got opinions, which is where it gets fun.



Corn

He does. He's not keen on what he calls the wild popularity of certain "shul microphones" — I take that as his word for the Shure SM7B, the one every podcast has — and the very deep proximity-effect sound everyone chases. He prefers brighter audio. But his principle is the interesting part: whether you're equipping or mastering your own voice or someone else's, you stay faithful to the voice. It's a canvas that rewards light touches. And he wants the light touches named concretely — mud reduction, de-essing, the like — with frequencies and the standard DAW tools.



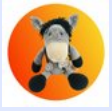
Herman

Which is a good question, because the answer is smaller than people think.



Corn

So let's start with the thing both fields share — the handful of moves.



Herman

Here's the claim I want to put on the table first. "Working with voice material" sounds like it covers a huge craft, but in practice it's a small, stable toolkit — a high-pass filter, a mud cut, maybe a presence lift, a de-esser, gentle compression — and the order you apply them matters more than which tools you use. EQ first, then compression, always. Compress a voice that's still full of rumble and the compressor reacts to the rumble — it's pumping on your desk bumps, not your words.



Corn

The compressor doesn't know what it's listening to.



Herman

It doesn't. It's just a level meter with opinions. And ACX, the audiobook platform, insists on the same sequence for mastering — clean the high and low ends before you limit, because a low hum otherwise drags the limiter around. Same logic, opposite end of the chain.



Corn

And the through-line for the episode, I think, is this: the same moves apply to human and synthetic voice, but on a human you're correcting a capture — the room, the mic, the plosives — while on TTS you're correcting the model's own artifacts. That asymmetry, not age, is why one field feels so much newer.



Herman

Which is a nice reframe, because the lazy answer is "TTS is newer, obviously." The lazy answer is wrong, or at least incomplete. But let's earn that later. The moves first.



Corn

The moves, with numbers. You're the one who's actually run a compressor that wasn't trying to kill him.



Herman

DJ years. Gain staging before breakfast. Right — the first move is the high-pass, or low cut. Eighty hertz, typical. Most speaking voices carry almost nothing useful below that, and the lows are where rumble lives — desk bumps, traffic, HVAC. A deep male voice may want seventy; a light voice can go up to a hundred. The test is your ears, not the number: raise the cutoff while listening, and stop just before the voice starts to sound thin.



Corn

Thin is the tell you've gone too far.



Herman

Professional practice tends to use a steep slope, twenty-four decibels per octave, set somewhere between eighty and a hundred and fifty — steep enough to kill the sub-bass garbage without stripping the body of the voice. A gentle slope lets rumble leak through for longer than you want.



Corn

Then mud.



Herman

Then mud. Two hundred to four hundred hertz, the low mids, where home recordings pile up energy. The sound of it is very recognizable — thick, cloudy, like talking into a cereal box. Every DAW has the same move for finding it. You take a bell band on your EQ, make it narrow, boost it hard — around plus ten decibels — and then drag that boosted band from about a hundred and fifty hertz up to five hundred while the voice plays.



Corn

And at some point it becomes unbearable.



Herman

Unbearable is the point. It'll suddenly sound awful and honky, and that honk is your problem frequency. Then you invert it: turn the plus ten into a cut of two to four decibels, medium width. The sources converge loosely on where mud lives — two hundred to five hundred depending on who you ask, Sound On Sound suggests a broad three-hundred-ish cut — but the sweep method matters more than the map, because every voice and room piles it up somewhere slightly different.



Corn

So the number is a starting guess and the sweep is the actual instrument.



Herman

That's the whole craft in one sentence, honestly. Next, presence. Consonants live around three to five kilohertz, and so does the sense that someone is talking to you rather than near you. A wide boost of two, maybe three decibels there makes words clearer, especially under music. More than about four turns harsh and tiring, and if the voice already cuts through, skip it. That's the discipline part — the move you don't do is also a move.



Corn

Presence is also the thing that bites you, isn't it. You lift four kilohertz and suddenly the S sounds stab.



Herman

You've just walked us into de-essing, and it's the one dynamic move in the chain. S, SH, the T in "tea" — sharp bursts somewhere around five to nine kilohertz, depending on the voice. Here's the misconception worth killing: de-essing is not an EQ cut at the sibilance frequency. A static cut there dulls every word all the time. A de-esser is a compressor keyed to that band — it only turns it down when an S actually happens.



Corn

A very small bouncer. Only works when the S shows up.



Herman

Finding the frequency is a nice parlor trick: play an S-heavy line — "six city sessions" is the canonical one — and watch the live spectrum, or let the tool's Analyze function report what it measures off the S sounds themselves. Then set it so the S sits back without turning into a lisp. If you hear "th" where an "s" should be, back off. That's the over-de-essing tell, and it's more common than the under-done version, because people set it once on headphones and never check on speakers.



Corn

Then compression. Numbers, Herman, I know you've been saving them.



Herman

Threshold around minus twenty decibels as a starting point, ratio two-to-one up to four-to-one for voice, attack five to twenty milliseconds, release a hundred to two hundred and fifty. Soft knee, for naturalness. But the number that actually matters is gain reduction — aim for three to six decibels of it on the loud words. Attack too fast flattens the snap of consonants; release too short and you hear the compressor flutter. If you can clearly hear it working — breathing, pumping, the room swelling up in the pauses — back it off.



Corn

The compressor should be felt, not heard.



Herman

There's a lovely framing I keep coming back to, from a training walkthrough by Jeff Pierson: used lightly, EQ and compression turn a decent recording into something that sounds like radio. Used heavily, they turn it into a fast-food drive-through speaker.



Corn

Which is the fate of roughly half the podcast section of the internet.



Herman

Pierson also has a worked example that ties the whole chain together, and I want to walk it because it shows how small the moves are. A council candidate, recorded close on a dynamic mic — boomy from proximity effect, jumpy levels. The fix: low cut at ninety hertz, which calms the boom. A sweep finds a honk at two hundred and eighty hertz, cut three decibels, medium width. Presence, plus two wide at four kilohertz. De-esser engaged because her S sounds now bite after the presence lift — notice how the moves cause each other.



Corn

The presence lift creates the sibilance problem.



Herman

It does, and that's why the order isn't arbitrary. Then the compressor, three-to-one, ten millisecond attack, hundred and fifty millisecond release, giving four to five decibels of reduction. His summary line: total change, measured, small. Change heard: she sounds like she is in the room with you, and nobody reaches for the volume knob.



Corn

That's the entire review a voice edit ever gets.



Herman

It's the only review that matters! Listeners never rate your EQ. They either reach for the knob or they don't.



Corn

And now, a very brief word from Corn's mother. Dorothy, are you there?

D

Dorothy

Hello dear, it's your mother. I've got one for you. Dorothy's History of Toast.

D

Dorothy

Mrs Beeton's Book of Household Management, from eighteen sixty-one, has a recipe for a toast sandwich. One slice of toast between two slices of bread, and it's recommended for invalids.

D

Dorothy

Well. You'd think toast would be the treat, and instead it's the filling. That's the history of toast for you, dear, mostly it's a bit of a letdown.

D

Dorothy

Anyway, I've got the kettle on and the washing out, so I'll let you go. Eat something proper, sweetheart. Bye bye now.



Herman

Delightful, as always. Where were we — the council candidate.



Corn

She was in the room with us. Now the two things Daniel really wants to argue about — the mic fashion, and why the synthetic side of this feels so much younger. The microphone first.



Herman

The SM7B. It's a good microphone — fifty years of professional radio and recording lineage, proper off-axis rejection, electromagnetic shielding, managed proximity effect. Nobody's claiming it's a fraud. But Daniel Miessler described its sound about as well as anyone has: pleasing, rounded, deep, but a little bit dead. It makes you sound authoritative.



Herman

And the trend he identified is that lots of podcasters, especially men, chase a super-deep sound because they believe deep reads as authoritative. His counter is empirical, not just taste: the extra bass sounds artificial over time, it's fatiguing, and it's hard to listen to in a car, because low rumble mixes with road noise.



Corn

The car argument is the strong one, I think. Taste arguments about deep versus bright go nowhere — you like what you like. But if most listening happens in cars, trains and kitchens, then the bass bump is optimizing for headphones in a treated room that most listeners don't have.



Herman

And the counter-model is sitting right there in public radio. The NPR sound — low bass, high clarity — comes largely from Neumann U87ai condensers. Condensers pull in much more range, way up toward twenty kilohertz, which makes them sound less dead and more realistic. The trade is that they capture everything — mouth noises, room sound, accurately and mercilessly.



Corn

So they're honest and the dynamic mic is flattering.



Herman

Flattering in a specific direction. Downward. And here's the thing I find funny about the whole fashion — the mic already has the fix built in. The SM7B ships with switches: flat, bass roll-off, and presence boost. Miessler's advice to SM7B users is to use those switches rather than pile bass on in post. His line is direct: resist the urge to add bass. It only sounds cool for a moment, and then is bad for listeners.



Corn

The microphone ships with an escape hatch and nobody pulls the handle.



Herman

His own chain is a nice portrait of the bright-but-not-tinny middle: bass cutoff below eighty, a slight lift in mid bass — note, a boost — a cut around fifteen hundred to remove bad mids, and a minor high shelf after seven thousand. Which also kills the folk wisdom that you should always cut and never boost. Gentle boosts add warmth and presence. Light touch doesn't mean subtractive only.



Corn

Good, because I was feeling slandered. I boost.



Herman

You boost constantly. You once argued for a high shelf on principle.



Corn

On principle and on evidence. So — the second half of Daniel's question. Why is voice engineering for TTS an infant field when human voice work has this decades-deep toolkit?



Herman

This is where the asymmetry matters. Human voice work corrects a capture — the room, the mic, plosives, proximity. The problems are physical and they've been the same problems since the microphone was invented. TTS work corrects the model's own artifacts: phantom silence, speaker drift, timbre that wobbles across segments. Different problems, so a different chain. A synthetic-voice post pipeline typically starts with voice activity detection cleanup — filtering silence and non-speech — and there's no human analogue for that. A human speaker doesn't emit phantom silence.



Corn

Our old voices did. I remember the gaps.



Herman

Documented, even. Then EQ and compression, then de-reverb and denoise, then limiting peaks to minus one and normalizing to around minus sixteen LUFS. Familiar tools, but bolted onto a front end that has no human equivalent. And the research literature confirms the split: it's all about controllability — voice design, timbre editing, prosody cloning — not post-hoc polish. CapTalk, out this spring, names the core tension as the conflict between stable timbre preservation and context-adaptive expression. EDICT, in October, points out that neither cloning nor text-based voice design lets you modify a reference's timbre, and proposes an edited acoustic reference to anchor voice identity across segments.



Corn

"Anchor voice identity across segments." That's faithfulness with a model number on it.



Herman

It's Daniel's canvas principle, stated as an engineering requirement. Change the sound, keep the person. MOSS-VoiceGenerator comes at it from the other end — studio-trained models are clean and well-articulated but lack the lived-in qualities of real human voices, so they train on cinematic content for real-world acoustic variation. The whole frontier is circling the same idea from three directions.



Corn

Which raises the question of whether this episode is one field or two. Same tools, different patients.



Herman

Well, corrective EQ on synthetic output differs from destructive EQ on human capture, in a specific way. On a human you're preserving something that was there — the room, the performance. On synthetic output there was no room and no performance. The old Chatterbox voices we retrained had no room tone, no natural dynamics. The EQ wasn't removing an accident of capture, because there was no capture.



Herman

And the human side has a formalized faithfulness contract that's worth spelling out, because it shows what "faithful" means when you write it down. ACX's audiobook standards: RMS between minus twenty-three and minus eighteen decibels, peaks no higher than minus three, noise floor no higher than minus sixty. And — this is my favorite requirement in all of audio — one to five seconds of room tone at the head and tail of every file.



Corn

Mandated silence.



Herman

Mandated continuity. So chapter one and chapter nine don't sound like different rooms. That's what staying faithful to a voice means when a platform has to enforce it at scale — the listener's sense that it's the same person in the same chair the whole way through. They even warn that Audible applies its own light dynamics processing after submission, so your levels shouldn't be pushed to either extreme.



Corn

You know what strikes me about that contract? We built the same one from the other direction. Room-tone beds at minus sixty in the gaps, because of Daniel's phone listening test. Two fields arriving at the same number for the same reason — the listener's sense that nobody left the room.



Herman

The specs converge even when the roads don't touch. So, Corn — the SM7B question, the one Daniel actually took a side on. Deep or bright. You've been sitting unusually quiet about it.



Corn

I'm on record. I said it earlier and I'll say it again, nobody reaches for the volume knob. The deep fashion fails that test in a car, and the car is where the knob lives.



Herman

You'd be a presence-boost man, then.



Corn

I'd take the U87 sound, if the U87 came with someone to clean the room. Sixty decibels of noise floor is a lot of housework.



Herman

It really is. Condenser honesty means you hear the refrigerator two rooms away.



Corn

Which is why the middle holds. Use the switches on the mic, resist the bass in post, and let the presence do the authority work. Deep says serious for the first ninety seconds and then says tired.



Hilbert

The serious sound is a reading voice, that's my issue with it. You hear that round, deep, dead tone and you know before the first sentence is out that the man is reading off a page. Talking has air in it. Herman keeps saying it his own way, but nobody's said the plain thing: nobody talks at a close dynamic mic like that unless he's been told to.



Corn

That's the tell I hadn't named. The proximity sound announces the script.



Hilbert

It does. And there's a de-esser question you two left swinging. You gave the range, five to nine kilohertz, and then said find it with a spectrum. Fine. But I'll tell you where it usually sits on a man reading closely — about six and a half, and it moves down as the day goes on, because the voice dries out by afternoon. I set it once for the morning and once for the afternoon, and the man I mixed for, a voice on a farming program, never knew I did it. He thought the mornings were just better takes.



Herman

Separate settings for morning and afternoon sibilance — because the voice dries out through the day.



Hilbert

Because the voice dries out. Corn says the moves cause each other — the presence lift creates the S problem. Same thing, in reverse. The day creates it too. I did jingles for a bread company for a while and the engineer there taught me the two-pass thing, and I never had a reason to drop it since.



Corn

The bread company engineer is now part of this episode's canon.



Hilbert

He had a cardigan and strong views about fifteen hundred hertz, same as your Miessler man. We agreed on that number and nothing else.



Herman

Six and a half kilohertz, set twice. That's the detail I'm going to be checking for on everything I hear this week.



Hilbert

Check it on an afternoon recording. The morning one will fool you.



Corn

It fits, you know. Miessler's whole chain — bright, present, the high shelf after seven — it's the sound of a man you believe is talking, not reading. The deep fashion isn't just fatiguing, it's telling on itself.



Herman

So the one thing to take from today, if you take one thing: the moves are small on purpose. A ninety-hertz cut, a three-decibel dip, two of presence, four of gain reduction. That's the whole difference between radio and a drive-through speaker, and it's the same handful of numbers whether the voice came from a larynx or a model.



Corn

And the faithfulness principle travels with it. Change the sound, keep the person — ACX writes it as room tone and RMS windows, EDICT writes it as an acoustic anchor, but it's the same sentence.



Herman

Which leaves us with the open question, the one I think is live: if the same handful of moves apply to human and synthetic voice, and the difference is what you're correcting, does TTS voice engineering converge on the same craft eventually — or does the model's own artifact profile keep it a separate discipline?



Corn

My guess is the second, for a while. The TTS chain will always have the VAD front end and the drift problem, and those don't shrink to zero.



Herman

And the research money is going the other way anyway — controllability, not post-hoc polish. Voice design, timbre editing, prosody cloning. The mastering side of TTS is staying practitioner-led, which is exactly where the new jobs are appearing. ElevenLabs is hiring audio engineering leads with human-edited dubbing experience. The old craft is being hired into the new field rather than replaced.



Corn

The people who learned the sweep method on humans are the ones being asked to do it to models.



Herman

They are, which is a nice thought to end on for anyone who spent years learning what two hundred and eighty hertz sounds like and wondered if it would ever matter again. It matters more than ever.



Corn

Producer credit where it's due: Hilbert Flumingtop, who has strong views about six and a half kilohertz and the cardigan of a bread-company engineer. If today was worth your ears, leave us a review, or find the whole archive at my weird prompts dot com.



Herman

If you want more of this, try episode ten, How ASR Went From Frustration To … Whisper Magic; and episode sixteen, On Deepfakes, SynthID, And AI Watermarking. This has been My Weird Prompts, the human-AI collaboration podcast.



Corn

See you tomorrow, and mind your mud.



Herman

Below minus sixty, always.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **Daniel Miessler, Analysis of Mics and Mic Sounds Used by Podcasters, published 2024-06-03, retrieved 2026-11-16.**
   https://danielmiessler.com/blog/podcast-audio

2. **Jeff Pierson, EQ and Compression on a Voice (Lesson 4 of 7), retrieved 2026-11-16.**
   https://jeffapierson.com/journal_tools/training/08-audio-production/04-eq-and-compression.html

3. **ACX / Scott Jacobi, Mastering Audiobooks with Alex the Audio Scientist, posted 2015-10-22, retrieved 2026-11-16. [primary**
   ACX is the platform setting the standard]
   https://www.acx.com/mp/blog/mastering-audiobooks-with-alex-the-audio-scientist

4. **Winberg, ACX check: audiobook RMS, peak, noise floor and room tone, requirements dated 2026-04-15, retrieved 2026-11-16.**
   https://winberg.ie/tools/acx-check/

5. **Trevor O'Hare, ACX Audio Requirements 2026: RMS, Peak & Noise Floor Specs, retrieved 2026-11-16.**
   https://www.trevorohare.com/blog/understanding-the-acx-submission-requirements-for-audio

6. **Craig Anderton, Sound On Sound, How To Make Your Vocals Twice As Good! Part 2, published August 2018, retrieved 2026-11-16.**
https://www.soundonsound.com/techniques/how-make-your-vocals-twice-good-part-2

7. **Finchley, Audio Engineering in a Professional Podcast Post-Production... The Process of Mixing, retrieved 2026-11-16.**
https://www.finchley.co.uk/finchley-learning/visual-podcast/audio-engineering-in-a-professional-podcast-post-production-concepts-and-practices-the-process-of-mixing

8. **Primer disparo, How to EQ vocals for a podcast, retrieved 2026-11-16.**
https://www.apovedam.com/en/free-podcast-course/eq-for-voice.html

9. **Mixing and Mastering AI, Voice Mastering: Podcast & Vocal Polish, retrieved 2026-11-16.**
https://mixingandmastering.ai/blog/voice-mastering-guide

10. **Baywood Audio, How to Mix Vocals in FL Studio (2026 Guide)**
Stock Plugins Only, retrieved 2026-11-16.
https://baywoodaudio.com/guides/how-to-mix-vocals-fl-studio

11. **SoundGuys, Shure SM7B review, retrieved 2026-11-16.**
https://www.soundguys.com/shure-sm7b-review-28438/

12. **Ray Ortega, The Podcasters' Studio ep. 103, THE SM7B**
Popularity, Noise, & First Mics, published 2019-01-15, retrieved 2026-11-16.
https://thepodcastersstudio.com/103-the-sm7b-popularity-noise-first-mics/

13. **HaoPicks, Shure SM7B Guide 2026, retrieved 2026-11-16.**
https://www.haopicks.com/shure-sm7b-broadcast-microphone-podcast-guide-2026

14. **Hacker News comment by roblh, 2023-11-27, retrieved 2026-11-16.**
https://news.ycombinator.com/item?id=38433348

15. **Voice Engineering and the TTS Recording Matrix, My Weird Prompts podcast on iVoox, retrieved 2026-11-16.**
https://www.ivoox.com/en/voice-engineering-and-the-tts-recording-matrix-audios-mp3_rf_183898102_1.html

16. **Zhang et al., TTSonic: A Unified Framework for Composable and Controllable Speech Generation, submitted 2026-11-03, retrieved 2026-11-16. [primary**
preprint]
https://arxiv.org/abs/2611.04282v1

17. **Su et al., CapTalk: Unified Voice Design for Single-Utterance and Dialogue Speech Generation, submitted 2026-04-09, retrieved 2026-11-16. [primary**
preprint]
https://arxiv.org/abs/2604.08363v1

18. **Zhao et al., Edit Who Speaks, Control How They Speak: Global Timbre Editing and Local Instruction Control for TTS (EDICT), submitted 2026-10-08, retrieved 2026-11-16. [primary**
preprint]
https://arxiv.org/abs/2610.11437v1

19. **Huang et al., MOSS-VoiceGenerator: Create Realistic Voices with Natural Language Descriptions, submitted 2026-03-30, retrieved 2026-11-16. [primary**

preprint]
https://arxiv.org/abs/2603.28086v1

20. **Hu et al., VoiceSculptor: Your Voice, Designed By You, submitted 2026-01-15, revised 2026-01-20, retrieved 2026-11-16. [primary**

preprint]
https://arxiv.org/abs/2601.10629v2

21. **StudyRaid, Understand audio Post-processing Integration Pipelines, retrieved 2026-11-16.**
https://app.studyraid.com/en/read/72764/3491441/audio-post-processing-integration-pipelines

22. **ElevenLabs Audio Engineering Lead posting via Jobgether, retrieved 2026-11-16. [primary**
employer posting]
https://jobgether.com/offer/6a79ee8c490731f1e1922810-audio-engineering-lead

23. **Dialpad AI Engineer, Agentic Voice (TTS) posting, closing 2026-08-31, retrieved 2026-11-16.**
https://www.jobs-cast.com/job/C6DN6cVlF/ai-engineer-agentic-voice-tts

24. **Harnham AI Voice Engineer posting, retrieved 2026-11-16.**
https://app.9am.works/job/4e67e344-4a39-4324-a3a9-d84919929203

25. **Tanveer Hussain, AI & Voice Engineer profile on Twine, retrieved 2026-11-16.**
https://www.twine.net/tanveer64