Episode Book

EPISODE #5851

The Earbud Assistant Problem Nobody Solves

PUBLISHED	11 October 2026 at 04:55 UTC
DURATION	23:45
LISTEN	myweirdprompts.com/episode/earbud-assistant-plumbing-problem-80605/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.3
SOURCES	14 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:40 UTC

SYNOPSIS

We've been promised a seamless voice assistant in our ear for a decade — and the model layer is finally good enough. So why does the thing still not work? This episode digs into the unglamorous plumbing that makes or breaks the wearable assistant: a hardcoded mic constant in Home Assistant's Android app that routes audio to your pocket instead of your earbud, a wake-word spotter that silently eats the first word of your command, and the brutal physics of fitting voice activation into a device with a four-kilobyte instruction cache. We look at Yandex's remarkable 200-kilobyte wake-word engineering, the volume-knob bug that took a month to find, and what Humane's $230M failure tells us about solving the hard problem while shipping a bad experience. The model isn't the bottleneck anymore — and that's a strange place to be.

TRANSCRIPT



Corn

Okay. I want to open by saying that I have been promised this exact thing — a little voice in my ear, always there, one word away — roughly once a year for a decade.



Herman

And how's it going?



Corn

I'm still reaching for my phone. Which is, I think, Daniel's whole point.



Herman

Go on then.



Corn

Daniel wrote in about the seamless assistant. The earbud, the worn mic, the wake word, the thing you just talk to. And his argument is that this already half-exists — you can pair a Bluetooth headset to an Android phone and say hey Google — but it keeps falling short, and it falls short on caveats. His phrase was "those fine technical details that seem like irrelevant facts but end up making or breaking the use case." The mic doesn't reliably capture. The wake word is brittle. The glue that lets the assistant actually do something isn't there yet. And then he asks the real question, which is: how far away are we, actually, from the thing becoming indispensable?

Herman

It's a good read.



Corn

It is. And the interesting part isn't whether he's right about the vision. It's that the model layer — the part everyone assumed was the hard bit — is basically solved. Which leaves us asking why the thing in your ear still doesn't work.



Herman

Then let's do the part nobody covers, which is the plumbing. Because you're right, the model is not the bottleneck anymore, and that's a strange place to be after ten years of assuming it would be.



Corn

So the vision is clear and the model is basically solved, which makes it worth asking why the thing in your ear still doesn't work.



Herman

Here's the framing I keep coming back to, and it's Yandex's, from their writeup on the Drops earbuds. For smart speakers, voice activation is a solved problem. You put a five-hundred-millilitre-ish battery in a mains-powered puck, a decent microphone array, a chip that doesn't care about power, and you can afford to listen forever. Earbuds are the opposite of that in every dimension. Tiny battery. Barely any memory. A chip with strict performance limits. Same problem, opposite physics.



Corn

And the form factor is the whole appeal, right? That's the thing nobody wants to give up. The value of a worn mic is that you don't have to take anything out of your pocket. Which means every compromise you make to fit it in your ear attacks the one property that made you want it.



Herman

And this category has a very expensive corpse sitting in the room, which is Humane. Two hundred and thirty million dollars raised. Six hundred and ninety-nine dollars for the pin, plus twenty-four dollars a month for the service. Discontinued in February of last year, and HP bought the IP for somewhere around a hundred and sixteen million.



Corn

The post-mortem on that one is what matters, though. Because the consensus wasn't "the AI wasn't good enough." The AI was reasonably sound. It failed on execution and on user experience. The thing did the clever part fine and the boring part badly.



Herman

Which is the thesis of this entire episode in one product. They solved the hard problem and shipped a bad experience. And the reason I bring Humane in is not to dunk on it, it's to set the stakes. There's a graveyard here, and the headstones don't say "the model was too small."



Corn

Right. So let's start with the baseline everyone assumes is fine, the Bluetooth headset and the phone in your pocket.



Herman

This is the one that gets me, because it's so small. Home Assistant's Android app had an issue filed in February of this year, number six four three three. And what it documents is: you pair a Bluetooth headset. You say the wake phrase. Audio output routes to the headset correctly — you hear the reply in your ear. But mic input stays on the phone. The app's recording logic is hardcoded to use the device microphone. There's no logic anywhere in it to check for and use the Bluetooth SCO path, because Android's default behaviour is to keep audio input on the phone's own mic unless the app specifically asks for the Bluetooth one.



Corn

So you're standing there, phone in your bag or your back pocket, talking to a headset that is listening through your trousers.



Herman

That's the issue's own words, more or less. It says this is a significant problem when the phone is not easily accessible — in a pocket or a bag — because commands are not understood or are missed entirely.



Corn

One line. One hardcoded constant.



Herman

One hardcoded constant, in one app, on one platform, and it deletes the entire reason the product exists. If the input silently falls back to the phone mic, then the earbud is decorative. You've built a hands-free assistant that requires your hands to be near your phone.



Corn

And notice what kind of failure that is. That's not a model failure. Nothing about that is a neural network being insufficiently clever. That's an app developer in a hurry writing the default and never testing the case that matters.



Herman

It also tells you something about incentives. The Bluetooth SCO path exists. Android can do it. The hardware can do it. Nobody wired it up, because wiring it up is work that doesn't demo well.



Corn

Now do the other half, which is the wake word.



Herman

Second Home Assistant issue, number six nine four three, end of May this year. And the report is: microWakeWord, which is the open-source on-device spotter they use, often misses the first word or words. So you say "Okay Nabu, what's the weather?" and it may not detect "what's the weather" at all. It just gets nothing. And "Okay Nabu, turn on the lights" comes through as "the lights."



Corn

It ate the verb.



Herman

It ate the verb, so now you've got a noun and no instruction, and the assistant either does nothing or does something confidently wrong. And this is not a fringe implementation. microWakeWord is what a lot of the local-first crowd runs. This is the state of the art for people who care enough to run their own.



Corn

So the brittleness isn't just that it misses the trigger. It's that the failure is partial and silent. It heard you. It just heard you wrong, and it has no way of telling you that.



Herman

There's an interesting asymmetry there, actually, because a total miss is fine — you say it again. A partial capture is worse than a miss, because the assistant acts on half a sentence.



Corn

And it doesn't ask. It doesn't say "the lights, what about them?"



Herman

No, because it doesn't know it's missing anything. That's the thing. A wake-word spotter doesn't produce a confidence score that's usefully calibrated. It fires or it doesn't. So "the lights" is a complete utterance as far as it's concerned.



Corn

Now — I want to go to the other side of this, because the wake-word engineering that does exist is impressive and I don't want to sell it short.



Herman

This is my favourite part of the whole research, Corn. Yandex's June writeup on fitting a spotter into about two hundred kilobytes.



Corn

Down from what?



Herman

Down from the roughly one point seven megabyte models they were running on smart speakers. So it's an order of magnitude smaller, and not by trimming — by rebuilding the architecture around the constraints. Their chip had two hundred and eight kilobytes of SRAM and a four-kilobyte instruction cache. Four kilobytes of instruction cache.



Corn

That's less memory than a single frame of video.



Herman

And the SDK imposed limits that shaped the whole design. Kernel size no larger than fifteen for standard convolutions, no larger than eleven for depthwise. No padding. No Hardswish activation. Those aren't preferences, they're hard walls, and the standard architecture you'd reach for first doesn't fit inside them. So they had to go to a deeper residual network instead.



Corn

So the tooling shaped the model, not the other way round.



Herman

The tooling absolutely shaped the model. And then they built a two-stage pipeline — a voice activity detector feeding a spotter — which gave them a five times reduction in system load, because the VAD is only active about fifteen percent of the time across seven and a half hours of indoor audio. So the expensive part only runs when there's a



Herman (continued)

human voice to run it on.



Corn

That's a nice trick. Cheap gate, expensive gate.



Herman

And a primary and secondary earbud scheme, where one bud does the listening and the other stays quiet, and they swap — that got them one point eight to two times the power savings. And all of it trained on about three hundred thousand recordings collected over three months, which they had to log in Opus compression, because uncompressed PCM logging was literally impossible at that volume. They couldn't store the audio fast enough.



Corn

Okay. And then the part I know you're getting to.



Herman

And then the bug. Which is my favourite story in the episode, because it's the thesis of the episode. They spent a month hunting a failure where the model appeared to be going deaf — the spotter just stopped responding. Turned out, during one of their logging runs, the recording levels were mismatched by a factor of ten, and that mismatch caused an integer overflow in the pipeline that broke the input scaling. So the model wasn't broken. The model was fine. The numbers going into it were wrong because a volume setting in a logging script was off by ten.



Corn

A volume knob broke the model.



Herman

And that is exactly what Daniel meant by an irrelevant fact that makes or breaks the use case. Nobody writes the paper on the volume knob. But the volume knob is why your product doesn't work.



Corn

And there's the false-trigger side, which I assume they also hit.



Herman

They had to disable the "next station" command, because subway announcements were triggering it. So people on the train were skipping tracks by accident, because a recorded voice said something in the ballpark of the trigger phrase.



Corn

Trains controlling my music.



Herman

And it's worth being precise about what that is, because it isn't a bug. False alarms and false rejections sit on opposite ends of one dial. You tune the threshold, and every notch you turn to catch more real wake words catches more subway announcements. It's not something you fix. It's something you choose.



Corn

So the honest state of play after all that.



Herman

There's a paper from MobiSys, called EarVoice. And they measured what premium earbuds actually do today — Pixel Buds, AirPods, Bose QC35. And all three require a tap or a button hold to activate the assistant.



Corn

So the most expensive consumer earbuds on the market are not hands-free.



Herman

Deliberately not hands-free. The tap is there to prevent accidental activation. The manufacturers looked at the false-trigger problem, and their answer was a finger.



Corn

They solved it by requiring the thing the whole product was supposed to eliminate.



Herman

Their numbers are worth having, too. Around ninety percent wake-word accuracy standing still, dropping to eighty-four percent while moving, and AirPods around ninety percent. And to get that, the research rig needed a little external dongle that cost about eight dollars and thirty cents — so the accuracy they achieved depended on hardware that isn't in the consumer product.



Corn

Because the mic in the consumer product is shaped by industrial design and not by the signal chain.



Herman

The mic is wherever the earbud had room.



Corn

So it can hear you, mostly. The harder question is whether it can do anything once it has.



Herman

Right, because a perfect wake word and perfect mic routing gets you a transcription. It doesn't get you an action. And that's the other half of the problem, and it's a different shape entirely. Because the number of assistants you might use is one thing, and the number of platforms they need to talk to is another, and the product of those two numbers is a mess. Every assistant times every platform is one integration, and each one has its own authentication, its own quirks, its own failure modes.



Corn

N times M.



Herman

And the plumbing problem from the first half of this episode generalises into exactly this. Same disease, different layer. The mic routing is one app forgetting to ask for the Bluetooth path. The integration glue is every app forgetting to talk to every other app.



Corn

So who's actually shipping something that takes action?



Herman

Plaud announced the One Explorer Edition in August. Two hundred and forty-nine dollars and ninety-nine cents, ships late September, 4G built in so it works standalone without a phone, and a thing they call the Plaud Agent that hooks into Gmail, Google Calendar, Notion, Slack.



Corn

And the tell?



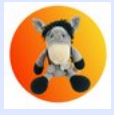
Herman

The tell is how you trigger it. It's an Agent Button. Not a wake word. You press a button.



Corn

Same as the AirPods. The premium product ships the finger.



Herman

And the coverage of it is unusually candid. It notes there are virtually no audio specifications published for it, because it's an AI product first and an earbud second. Which tells you the company knows exactly which half of the problem it's competing on.



Corn

They've decided the audio is a solved commodity and the agent is the product.



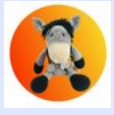
Herman

Which is a defensible bet, and also a bet that the plumbing underneath them will hold. Which is the thing nobody controls.



Corn

Now — the standard that's supposed to make this go away.



Herman

MCP. Model Context Protocol. It's the emerging answer to N times M: instead of writing one integration per assistant per platform, you write one MCP server per platform, and every assistant that speaks MCP can use it. And it does help. But it's not free, and the costs are exactly the sort of fine detail Daniel was pointing at.



Corn

What kind?



Herman

There's a benchmark paper, LiveMCPBench, that notes a large gap between how MCP is actually used in the wild and what the current evaluations test. So the standard exists and the measurement of the standard lags behind it. And one analysis is blunt that MCP doesn't make every tool safe — the protocol describes how a tool is called, not whether calling it is a good idea. And it points out that generic MCP servers cost context, reduce tool-selection accuracy, and add an authentication and security surface you don't control.



Corn

So the glue you outsourced comes back as a new attack surface.



Herman

And as a context budget problem. Every tool you expose eats room in the model's window and gives it one more way to pick wrong. There's a sweet spot in how many tools you hand an agent, and it's not "all of them."



Corn

Which is a funny place to land, because the whole pitch of MCP was fewer moving parts.



Herman

Fewer integrations, yes. Not fewer problems. You consolidate the plumbing into one place and then you have to secure and tune that one place.



Corn

Then there's the hardware wave, which is arriving whether or not the software is ready.



Herman

Qualcomm announced Snapdragon Sound Elite Gen Two at the Snapdragon Summit at the end of September. One hundred and twenty-eight billion operations per second, up from sixty-four billion on the previous generation. Forty percent less power. Thirty percent smaller. And it runs apps on-device without a phone at all.



Corn

So the chip is no longer the excuse either.



Herman

The chip is not the excuse. And Andrew Laister from Qualcomm gave the line that I think will be quoted for a year, which is that it's like the first smartphone moment with the Play Store — except that app store for your earbuds is currently owned by the device maker.



Corn

Which is a lovely quote and also a warning label.



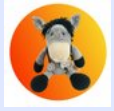
Herman

It's a warning label. Because if the earbud app store is owned by whoever made the earbud, then you've replicated the N times M problem one layer down. Your assistant has to be written for Samsung's store, or Qualcomm's, or whoever's, and now the fragmentation isn't between platforms, it's between pieces of hardware.



Corn

So the standard that would solve it — a common runtime for earbud apps — doesn't exist, and the party who'd have to give it up is the party currently holding it.



Herman

Laister concedes the other half of the problem himself. He says there will likely be an impact on battery drain, because these devices are designed to be used continuously. Always-on is the feature. Always-on is also the power draw.



Corn

You can hear the marketing meeting. "Always listening" is the pitch and "always draining" is the spec sheet.



Herman

The privacy angle sits right next to the feature angle, because they're the same thing described from two directions. The chip harvests audio and camera data to build what they call a personal profile. Contextual awareness and surveillance are not two products. They're one capability with two press releases. And the only reassurance offered on the camera side is that it's low resolution and AI-masked.



Corn

Low resolution and masked. That's the whole guarantee. So let's answer Daniel's question directly, because we've been circling it. How far away, actually?



Herman

Hardware is arriving fast. Yandex shipped a working two-hundred-kilobyte spotter. Qualcomm doubled the AI throughput. Samsung launched Galaxy Buds On at the start of October — clip-on open-ear, nine and a half hours of battery, improved voice pickup. The pieces are all landing.



Corn

And yet?



Herman

Yet I could not find a single standalone, shipping, truly hands-free wake-word earbud assistant that works reliably without a phone and takes robust actions. The closest thing is Plaud, and it's button-triggered. The other closest thing is Qualcomm's chip, which is announced, with no devices and no timeline.



Corn

Not one.



Herman

That's the answer to the question. The technology is close. The product isn't.



Corn

Name the hurdles, then, because that's the other thing he asked.



Herman

There are four, and they're different kinds of problem, which is why they haven't been solved together. Reliable Bluetooth mic routing is an operating-system and application-level gap. The false-alarm versus false-rejection tradeoff is a tuning problem you can't win, only balance. Environmental false triggers are a consequence of the balance you pick. And robust agentic action-taking is the N times M glue problem that MCP helps with but doesn't close.



Corn

None of the four are model problems.



Herman

None of them. Every single one is plumbing.



Corn

Which is why the fix isn't coming from the place you'd expect. Because the plumbing is owned by the device maker, the operating system, and the platforms — and not one of those three owns the seamless experience. The device maker wants you in their app store. The OS vendor wants you using their assistant. The platforms want you inside their app. Nobody gets paid for the handoff that makes the whole thing disappear.



Herman

That's when the plumbing problem turns into an incentive problem.



Corn

Say that again, because I think that's the real finding.



Herman

The failure isn't that nobody can build this. The failure is that the person who'd have to build it doesn't own the parts, and everyone who owns a part is fine with the current state. You'd need one company to own enough of the stack — the chip, the OS, the assistant, and the integrations — that the fine details stop leaking out. And the only companies that own that much of the stack today are also the ones with the least reason to open it up.



Hilbert

Two hundred and eight kilobytes.



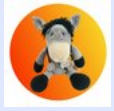
Corn

What?



Hilbert

The Yandex number. It's two hundred and eight kilobytes of SRAM, not two hundred. You said the chip had two hundred. It had two hundred and eight.



Herman

He's right.



Corn

Of course he is.



Hilbert

I did a stint doing QA listening for a voice product. Weeks of it. Sitting in a room saying the same phrase into a device, thousands of times, while somebody watched a screen. And the thing that killed that product was not a missed wake word. Not once was it a missed wake word. It was that the device worked for me and failed for everyone else in the building.



Corn

Why?



Hilbert

My voice sat in the band the model was trained on. So every test I ran came back clean, and the team shipped it, and it fell over the moment it met a real customer. That's what you never test — the thing the way a person uses it.



Herman

They hired you for the voice.



Hilbert

I was statistically average. That was the qualification. They kept a roster of people with average voices on retainer. They called it The Normals.



Corn

The Normals.



Hilbert

They paid us in gift cards to a smoothie chain. Twenty dollars a session. I did about a hundred and forty sessions before my voice drifted after a bad cold, and they let me go.



Corn

Your voice drifted.



Hilbert

Colds change your voice. Permanently, sometimes. I came back from three weeks off and they ran the measurement and said I was outside the band. That was it. Eleven months of that job.



Herman

And the roster?



Hilbert

Still on the payroll system as far as I know. I still get the gift cards. It's been a long time and they still show up. Twenty dollars a month, automatic, from a company I haven't spoken to since.



Corn

You've just never mentioned it.



Hilbert

I'm not calling them. They might want the smoothies back.



Herman

How many smoothies are we talking about?



Hilbert

A lot of smoothies, Herman. And they discontinued the flavour I liked, so now I've got a stack of cards I can't spend and a moral obligation not to ask why.



Corn

The roster was called The Normals and it's still paying you.



Hilbert

It's still paying me. That's the entire business. The model was fine, the room was fine, the device was fine — and it worked for the one person who happened to sit in the middle of the distribution, which was the person they'd hired to test it.



Herman

That's the false-rejection problem as a staffing decision.



Hilbert

That's what it was. They had a number for the average voice and they hired the number instead of a hundred different people. Then they shipped.



Corn

The smoothie chain discontinued your flavour.



Hilbert

A while back, yeah.



Corn

Which takes us back to the question Daniel actually asked. How far away is this, really?



Herman

The first company to make this indispensable probably won't be the one with the best model. It'll be the one that owns enough of the stack to make the fine details disappear — the mic routing, the mic array, the wake-word threshold, the integrations — all in one place, so the user never meets any of them.



Corn

The odds of that being a company that plays well with everyone else's stack?



Herman

Low.



Corn

That's the actual answer, I think. Not "three years" or "five years." It's that the thing is technically available now and commercially nobody wants to build it.



Herman

Most people think the bottleneck is the AI. The model gets the blame because it's the visible part.



Corn

The model is the one part that's already done. The thing that's missing is a hardcoded line in an app nobody's incentivised to fix.



Herman

Thanks as always to Hilbert Flumingtop, who produces this show and who is, apparently, still on somebody's payroll somewhere.



Corn

If you enjoyed this one, try episode ten, How ASR Went From Frustration To … Whisper Magic; episode five, Fine-Tuning ASR For Maximal Usability; and episode seven, Building Custom ASR Tools. This has been My Weird Prompts. If you want to send us one of your own, it's the Telegram bot — t dot me slash MWP listener bot.



Herman

We'll be back soon.

SOURCES

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

1. **Yandex, 2026-06-12 [primary]**
   https://medium.com/yandex/how-we-reimagined-voice-activation-for-yandex-drops-and-fit-an-entire-model-into-200-kb-70bbf1c2f213

2. **Home Assistant Android, 2026-02-12 [primary]**
   https://github.com/home-assistant/android/issues/6433

3. **Home Assistant Android, 2026-05-31 [primary]**
   https://github.com/home-assistant/android/issues/6943

4. **EarVoice, MobiSys '24 [primary]**
   https://dl.acm.org/doi/epdf/10.1145/3643832.3661890

5. **Live Science, 2026-09-25**
   https://www.livescience.com/technology/electronics/new-kind-of-ai-audio-chip-transforms-earbuds-into-always-on-intelligent-devices-that-run-without-a-smartphone-how-does-it-work

6. **Android Authority, 2026-08-27**
   https://www.androidauthority.com/plaud-ai-earbuds-take-notes-action-3701242/

7. **Tech Times, 2026-10-01**
https://www.techtimes.com/articles/328365/20261001/samsung-galaxy-buds-are-here-new-design-95-hour-battery-galaxy-ai-features.htm

8. **The Silicon Report**
https://www.thesiliconreport.com/articles/rabbit-r1-humane-ai-pin-hardware-postmortem-edge-cellular-teardown

9. **Layer3 Labs, 2026**
https://www.layer3labs.io/comparisons/rabbit-r1-vs-humane-ai-pin

10. **appearan.com, 2026-01**
https://www.appearan.com/2026/01/the-ai-shift-no-one-is-talking-about.html

11. **LiveMCPBench, KDD**
https://dl.acm.org/doi/abs/10.1145/3770855.3817478

12. **Nango Blog**
https://nango.dev/blog/mcp-vs-tool-calls-for-ai-agents/

13. **prodSens.live, 2026-07-15**
https://prodsens.live/2026/07/15/building-a-production-mcp-server-how-we-made-goodbarber-agent-ready-without-the-glue-code/

14. **Hacker News, Hacking the Humane AI Pin, 2025-10-08**
https://news.ycombinator.com/item?id=45515915