Episode Book

EPISODE #6546

The Wake-Word Race Inside Home Assistant Satellites

PUBLISHED	13 November 2026 at 12:16 UTC
DURATION	27:44
LISTEN	myweirdprompts.com/episode/home-assistant-voice-satellite-race-39638/
HOSTS	Corn, Herman, Clive, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.26
SOURCES	11 cited by the research (listed at the end)
BOOK GENERATED	14 November 2026 at 06:47 UTC

SYNOPSIS

Multi-room Home Assistant voice setups rely on cheap satellites—like the Voice Preview Edition or ESP32-S3-BOX-3—streaming audio to one central hub. Because wake-word arbitration is a pure network race with no volume comparison, the farthest satellite can win and silently drop your command. We unpack the real-world pitfalls of multi-room voice: why turning down sensitivity backfires, how power-saving modes skew packet timing, why duplicate suppression fails, and what happens when a satellite triggers on a TV or its own voice.

TRANSCRIPT



Corn

Two satellites hear you say the wake word. The one that wins is whichever one happened to be at a good point in its CPU loop. Distance is not a factor.



Herman

I love that you opened with that, because that's the whole episode in one sentence and most people will not believe it.



Corn

Belief isn't required. Packets are.



Herman

Daniel's prompt this time came out of a visit. A friend of his came over who's all-in on Alexa — Echo in the bathroom, the bedroom, the living room. Daniel doesn't use Alexa, but the setup got him wondering what the same thing looks like on Home Assistant, and he suspects, correctly, that people don't actually put a Raspberry Pi in every room. What he wants is the part nobody writes on the box: the non-obvious problems. If he speaks in one room, how do you stop the satellite next door hearing it and acting? How do you stop one satellite's speaker being picked up by its own microphone, or its neighbour's? How do people set this up so it actually works, and what are the pitfalls nobody warns you about?



Corn

Fair warning to Daniel's friend: the answer to most of those is "badly."



Herman

Mostly badly, yes. But let's do this in order, because the first misconception is the funniest one.



Corn

The Pi in every room.



Herman

Nobody does that. The topology everybody actually builds is one Home Assistant instance running the whole voice pipeline — wake word, speech-to-text, intent handling, text-to-speech — and then each room gets what's called a satellite. A satellite is a cheap microphone-and-speaker endpoint, deliberately dumb. It streams audio to Home Assistant and plays back whatever Home Assistant sends it. It doesn't think. It's an input-output device with delusions of grandeur.



Corn

A headphone that believes in itself.



Herman

And the hardware is cheap. The official Home Assistant Voice Preview Edition is about fifty-nine dollars. The ESP32-S3-BOX-3 is around forty-five, and you can use its onboard speaker for music, not just responses. Or the real DIY path: a Pi Zero 2 W with a ReSpeaker two-mic HAT, about thirty-eight dollars in parts, at the cost of an evening of soldering and case-fitting.



Corn

An evening of soldering to save eleven dollars.



Herman

Some people enjoy the soldering, Corn.



Corn

Some people enjoy the baffle board too. I don't judge.



Herman

The whole five-room build — kitchen, living room, office, bedroom, garage — totals around two hundred forty-six dollars in satellite hardware, and under four hundred for a whole-house setup including music zones. Which sounds great until you remember Echo Dots retail around fifty dollars and go on sale for twenty or twenty-five constantly. So you are not saving money. You're buying capability and control, not a discount.



Corn

Which is the most honest sentence in the smart home industry.



Herman

And the guide everyone points to is honest about the other thing too, the thing Daniel's friend should hear before buying anything: a single Voice PE in the kitchen plus a Music Assistant zone on an existing speaker covers ninety percent of households. Multi-room earns its complexity only if multiple people want audio in different rooms at the same time, or the footprint is large, or you're deliberately retiring an Alexa or Google fleet.



Corn

Three Echos, in a flat.



Herman

Three Echos in a flat is a lifestyle, not a requirement. But Daniel's asking about the network of satellites, and that's the interesting episode, so let's take the premise seriously. One more setup detail, because it trips people up: the satellite's own speaker and the room's music zone don't have to be the same device, and in most real builds they shouldn't be. You route "play jazz in here" to a Voice PE's one-inch driver and you get jazz through a speaker that sounds like a transistor radio drowning. The correct pattern is two entities per room — the assist satellite and a media player — both assigned to the same area. Home Assistant doesn't need them to be the same box. It needs them to agree on which room they're in.



Corn

So the hardware is solved, the topology is solved, the prices are known. The interesting problems are all in the software layer, and every one of them starts with the same question: which satellite did you actually talk to?



Herman

Michael Hansen — synesthesiam, the creator of Rhasspy, Piper and Wyoming, so the person who built most of this stack — put it flatly in a GitHub issue about duplicate wake-ups. Wake word detection is done locally on each device, and there's a chance it will activate faster on the device farther away, because of where the CPU happens to be in its loop. On the Home Assistant side, they just listen to whatever satellite starts a pipeline first. If two try to start one with the same wake word, the second gets denied.



Corn

Read that back to yourself slowly. The system for deciding which room you're in is a race.



Herman

A race with no lane assignments. No volume comparison, no signal strength, nothing. Whoever's packet arrives first, wins. And that produces the nasty part, which is that the failure is silent. There's a community thread from August by a user named Linwood Ferguson that documents this about as well as anything I've read. He has three Voice PEs — office, living room, kitchen. Living room and kitchen are maybe five or six meters apart, office is twelve meters down a long hall. When the wrong satellite wins the race, the far one hears attenuated, distant audio — and here's the key — the speech-to-text can't transcribe it. His words: the kitchen says no text and nothing happens.



Corn

Not "the wrong room answered."



Herman

No. Nothing happened at all. The office logs a duplicate, the living room logs a duplicate, the kitchen logs no text, and you're standing there repeating yourself to a house that has already decided, incorrectly, that it heard you fine.



Corn

A reliability cliff, not an annoyance. With Alexa, the wrong room answering is a comedy. Here it's an outage.



Herman

And Ferguson did the work, this is what I love about the thread. He didn't just complain. He ran packet traces and found that arrival time at Home Assistant was definitely the issue — Home Assistant is picking the right one in terms of which arrives first, it's the actual packet arrival times that come out of distance order. So the arbitration logic isn't broken. The network is reordering the race while the logic faithfully rewards the reorder.



Corn

The referee is fine. The track is on a hill.



Herman

And his Wi-Fi was excellent — minus forty-four to minus forty-six dBm at both the access points and the devices, five APs, low airtime utilization. This is not a man with a weak signal. This is a man whose house is faster than his arbitration.



Corn

What did he actually change?



Herman

The sensitivity dial, first, and it's a lose-lose. At the lowest useful setting, slightly sensitive, a quiet voice a meter away may fail to trigger at all. Turn it up and the cross-room problem gets worse. The setting that stops the neighbour's satellite hearing you also stops you hearing you.



Corn

A dial with no winning position.



Herman

Then he found something better. The Voice PE ships with power saving set to light — despite the device not being battery powered, which he notes with what I can only call justified bafflement. He turned it off, and in ping testing it made a big difference; in brief



Herman (continued)

testing after that, he couldn't reproduce the problem. I'd stress that no one else in the thread confirmed it, so whether it's a real fix or coincidence is open.



Corn

"The device was sleeping through the job it was hired for" is a very Home Assistant sentence.



Herman

There's one more layer, and it's the one Daniel specifically asked about — how do you stop the next room acting on your command. Sometimes you don't. Ferguson reports that duplicate suppression simply fails once or twice a day: two satellites hear the wake word, both process the whole request, both respond. His reading of the code is that the duplicate-suppression window is two seconds, and he argues it should be half a second to one, or that the first arrival should be held briefly while gain is compared.



Corn

Two seconds is an eternity in network terms.



Herman

It's enormous. And his remark about it is the one that stays with me — generally the slower duplicate is a no-op, something like turning on a light or opening a door, but eventually he's going to run into something that does not play well being done twice.



Corn

My garage door would like a word.



Herman

Yours, mine, and anyone's whose automation set includes anything with volume or heat. Now — the fix that should exist, doesn't. A community member proposed it in that same GitHub issue in December of twenty twenty-five: carry the audio intensity at the moment



Herman (continued)

of wake-word detection, have Home Assistant wait about a hundred milliseconds for competing activations, pick the loudest, maybe make first-to-arrive versus highest-intensity a config option. Someone else replied, and I'm quoting because the typo is doing honest work, "Good Idea. I am struggeling with the exact same issue here."



Corn

Struggeling. Verb.



Herman

Verb. And the issue sat there until August of this year, when it was closed as stale. Never implemented. The maintainer's position from the start was that there might be nothing to do with the current hardware and software stack, which is at least an honest engineering answer. The devices don't report how loud you were, so the hub can't compare loudness, so it doesn't try.



Corn

The satellites know where they are. They just don't know where you are.



Herman

They don't know they're in a house with other satellites. Each one is running its own little local race and the hub just picks the first one over the line. Corn, let me ask you something, because you'd be the user here and I want your read. Given that the wrong satellite winning means silent failure — would you rather the far one won and got nothing, or would you rather have both respond every time? Because the duplicate failure at least tells you something happened.



Corn

Both, honestly. Wrong-but-loud beats right-but-mute. If the kitchen answers, I know the kitchen heard me, and I can walk two meters. If nothing answers, I'm debugging a house.



Herman

That's the diagnosis most of the thread eventually reaches too — people would trade annoyance for feedback. The really frustrating cases are the ones where you can't tell whether you mumbled or the network mumbled.



Corn

And the TV? Everyone with a living room satellite has a TV in that living room.



Herman

Separate failure mode, and no wake-word tuning fixes it. A satellite in a room with a television will false-trigger on dialogue, full stop. One user in that cross-room thread reported trying everything to separate his living room and kitchen acoustically, nothing worked, and the living room box also false-triggered from the TV — so he shut the living room box down and uses Google there while he keeps testing the kitchen one.



Corn

The concession pattern. Every multi-satellite house has a room that's surrendered.



Herman

There's a line in the Big Iron guide I keep coming back to, on cross-room bleed: lower the sensitivity on the satellite nearer the noise source, or physically reposition, because there's no purely software fix for two microphones both hearing the same voice clearly.



Corn

That's the arbitration problem, then — the wrong satellite wins and the command silently dies. But there's a second problem that's arguably funnier, and it's the one where the satellite hears itself.



Herman

Oh, this one. GitHub issue on the Voice PE firmware, still open as of late October. "Voice PE triggers itself while speaking." The reported example uses a custom wake word, Jarvis, and I want to give it to you as a transcript because the transcript is perfect. The user says, "Jarvis, tell me something about yourself." Jarvis answers, "My name is Jarvis, I am your personal assistant." Now — Jarvis just said its own name out loud. Its own name is the wake word. So it hears the wake word, starts listening while it is still speaking, catches the tail of its own sentence — "I am your personal assistant" — treats that as the user's command, and answers it.



Corn

What did it say to that?



Herman

"That's nice of you, but actually I don't require a personal assistant."



Corn

A machine that doesn't need an assistant. It's more self-sufficient than I am.



Herman

A commenter using the stock Okay Nabu wake word confirmed it's not limited to saying the wake word itself — the device hears its own response and replies to the snippet of its own response that it heard, and it begins talking to itself until it reaches a point where it doesn't expect a follow-up from the user. Another user reports it feeding back three or four times before stopping, and — this is the detail that should be on a poster — it started happening after he offloaded faster-whisper to a separate machine.



Corn

The optimization made it worse.



Herman

Faster transcription means faster turnaround, which widens the overlap window. And a contributor on that thread found a root cause in the state machine that rhymes with it: the end-of-speech-playback transition was triggered when the streaming stopped, not when the playback stopped. The device would move on to the next state while the TTS audio was still coming out of the speaker. That easily happens if the stream is significantly faster than real time.



Corn

So a fast TTS server can outrun the speaker it's feeding.



Herman

Physically outrun it. The pipeline finishes delivering before the sound finishes happening. I want to say that's a unique bug, but honestly it's the whole genre — every component in this stack is racing every other component, and the race only works when the sound is the slowest thing in the house.



Corn

Which it usually is. Until you buy an NUC.



Herman

And then your own house is too fast for itself. There's also a maintainer-side note in that thread worth flagging for anyone tempted by Jarvis-style custom wake words: the community-generated ones, most of which are built on Piper synthesis, are of poor quality with a large number of false positives. So some of this is people picking wake words that are acoustically soft targets.



Corn

"Jarvis" is two open vowels and a wish.



Herman

It's a lovely word that means nothing acoustically. So the workaround people converge on — and I want you to hear what it costs — is muting the microphone during playback. One user built an automation that latches a mute whenever playback and listening are detected at the same time, and unmutes when playback ends. He calls it less than ideal, because you cannot interrupt during playback. Without negative feedback filtering of the audio, he says he can't think of a better solution than cutting off the mic.



Corn

He's right, and that's the sad part. There isn't a better simple one.



Herman

So now we're at the design triangle, and this is the real answer to Daniel's question about a satellite's speaker being picked up by its own or another's microphone. Three corners. Corner one: real acoustic echo cancellation, AEC. The device keeps a copy of exactly what its speaker is playing and mathematically subtracts it from what the microphone hears, leaving only other sounds. That's also what makes barge-in possible — you can talk over the assistant because it can tell your voice from its own. Corner two: mute during playback. Easy, reliable, and it kills barge-in. Corner three: do nothing and accept self-triggering.



Corn

Pick your favourite way to be interrupted.



Herman

There's a project called wyoming-echo, about four hundred sixty lines of Python, that sits in corner two and is very upfront about it. It pauses the satellite during text-to-speech playback, computes the exact playback duration from the raw audio byte stream — sample rate, bit width, channels — waits that long plus a five-hundred-millisecond safety margin, then resumes. Its README's known-limitations section says, verbatim, no barge-in, you can't interrupt the assistant mid-response.



Corn

The five-hundred-millisecond safety margin is my favourite character in this story. Polite to a fault.



Herman

And the reason it's needed is the same race again — the computed duration and the actual speaker disagree by a few hundred milliseconds, and if you unmute early you catch the tail of your own voice. Whereas corner one is getting real: the Ava project shipped software AEC in release point four point eight — it knows exactly what it's currently playing and subtracts that reference signal from the microphone input, so wake-word detection and streaming stay robust even while music or responses are playing. That's the direction the whole stack has to go.



Corn

We touched this in the episode on building custom ASR tools from scratch, by the way — the speech-to-text stage of the pipeline being the part that punishes you for everything upstream. Same lesson, louder.



Herman

Same lesson, and the numbers explain why. Official Home Assistant docs put Whisper transcription at around eight seconds on a Raspberry Pi 4 versus under a second on an Intel NUC. Eight seconds of house where anything can happen to a session. And Piper generates about one point six seconds of speech per second on a Pi with medium models — so on modest hardware the sound is comfortably the slowest thing in the house, which is precisely why the self-triggering shows up when people speed the servers up. The bug is the luxury model.



Corn

And now, our regular segment. One-Star Getaways.

C

Clive

This is One-Star Getaways. In Penang I stayed at the Sea Breeze Heritage Annex, and the sea breeze was a laminated photograph of a sea breeze taped over the air conditioning vent, which hummed like a fridge full of bees. The room service menu offered a view of the harbour for an extra twelve ringgit, and the view was the car park with a picture of a

C

Clive (continued)

boat stapled to the fence. One star.

C

Clive

I recommend it for anyone looking for a disappointing trip this winter, and tell them Clive sent you, they will not know who that is.



Corn

Twelve ringgit for a car park. That's the satellite-arbitrage of tourism.



Herman

You're halfway to booking it, I can hear it.



Corn

The photograph of a sea breeze is at least quiet. Back to the mute problem — there's a layer of mitigations below the big three, right? What do people actually run day to day?



Herman

A menu, and every item on it has a price. Different wake words per room — Home Assistant supports swapping the on-device model, hey Jarvis in one room, Okay Nabu or Alexa in another, via a firmware substitution. Works. The cost is you have to remember which room you're in.



Corn

If I could remember which room I'm in, I'd have fewer problems generally.



Herman

Mute unused rooms — the man who surrendered his living room box is the honest version of that. Lower sensitivity nearest the noise source, at the cost of quiet speech failing. Disable power saving, per Ferguson. Then the unglamorous one that fixes more than any of them: assign areas properly. The Big Iron guide calls missing area assignment the single most common reason a setup fails in the field but works fine in the demo video.



Corn

The demo video. Filmed in a room with one satellite and no television.



Herman

Filmed by a person who has never had a kitchen argument carry into a hallway. There's also a glue layer worth knowing about — the assist satellite states in Home Assistant, listening, processing, responding, idle, with matching conditions. The official docs include a ready-made automation that drops a media player to twenty percent volume when a satellite starts responding and restores it on idle. And one that alerts you if any satellite has been stuck responding for thirty seconds or more — to catch satellites stuck in the responding state that silently waste resources.



Corn

There's a metaphor in a stuck-responding detector. I'll leave it where it fell.



Herman

Best place for it. Corn, can I ask you the comparison question, because you'll answer it better than I will. Alexa and Google do multi-room. The friend has three Echos and it works. So the obvious read is that the closed ecosystems solved this and the open one hasn't. Is that the right read?



Corn

No, and Ferguson's thread has exactly the line for it. Google has been known to fail to detect duplication too, and frequently answers from the wrong satellite. The difference is that when Google hears from a far satellite, it also understands. Their far-field speech-to-text is good enough that a distant, attenuated command still transcribes. So the wrong room answers, and it's annoying, and you move on. In the Home Assistant pipeline, the wake word may be heard but the STT isn't understood, because of distance,



Corn (continued)

attenuation, ambient noise. Same problem. They're not solving arbitration — they're burying it under a better transcription engine.



Herman

Which reframes the whole comparison. The closed ecosystems are hiding a hard problem behind better speech-to-text, not solving it.



Corn

The Echo Dot is a twenty-five dollar speaker with a transcription engine doing the work of an arbitration layer it doesn't have.



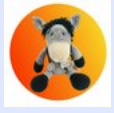
Herman

And there's a line in the Big Iron guide that completes it — every closed voice ecosystem gets multi-room audio right on day one because the vendor controls the speaker, the app, the streaming backend and the room list. The moment you decide you'd rather not have a single company deciding which streaming services survive next quarter, you inherit the job that vendor was doing for free.



Hilbert

Nakamichi LM-1, forty pounds, the pawn shop on Wharf Road, still in the box.



Herman

That's a specific start.



Hilbert

I'm getting there. I ran two of those voice satellites, one in the kitchen, one in the front room, for most of a year. Not the Home Assistant ones. Earlier. Worse. The front room one kept hearing the kitchen, the kitchen one kept hearing the front room, and both of them heard the radio, which was the Nakamichi's whole job. My brother-in-law does AV installs for hotels and he says the satellites are all fine and the problem is always the wall — but



Hilbert (continued)

don't trust him, he still owes me for a parking ticket. I asked the council's building records office what the original dividing wall was made of, to settle whether it was plasterboard or block, and a Mr. Halsey there took five months to send me one photocopied page. It was block. The photocopy was sideways.



Corn

So what did you do with the block wall?



Hilbert

Nothing to the wall. I put a sheet of acoustic foam floor to ceiling in the doorway gap where the old serving hatch used to be. That's the only thing that stopped the cross-talk, and it stopped it completely, which is more than any setting on either device managed. My claws made a mess of the fitting, took most of an afternoon. When I dismantled the voice system last spring, I left the foam where it was. It's still there. Herman, you spent forty minutes this afternoon on a race between two packets, and it's the first time it's occurred to me the system was never broken. The foam did more work than the firmware ever did. Whoever fitted that serving hatch, they did a good job, and it's a shame to cover it.



Herman

The foam is still up.



Hilbert

It blocks a draught.



Corn

The draught defence. Legally unassailable.



Herman

You know, the physical layer point is real — that cross-room thread had a man trying everything to separate two rooms acoustically and nothing worked, because the rooms were open to each other. The people who fix it fix it with furniture, bookshelves, walls. A doorway is a microphone cable between rooms.



Corn

A hundred-millisecond wait-and-compare-loudest fix, and the winning hardware is a sheet of foam and a grudge against a photocopy.



Herman

The misconception to leave people with, then. The one I'd want Daniel's friend to walk away without: that the satellite nearest you wins the wake word.



Corn

It doesn't. Wake word detection is local and asynchronous, and Home Assistant takes whichever satellite starts a pipeline first — which can be the device farther away because its CPU loop was in a good spot, or its packets arrived first. No loudness, no distance, nothing.



Herman

And the follow-on mistake is thinking the wrong satellite answering is the bad outcome. The bad outcome is silence — the far one wins, can't transcribe you, and nothing happens at all. You're not annoyed, you're unsupported.



Corn

The open question that matters is whether loudest-mic arbitration ever lands in Home Assistant. The proposal sat in that issue for eight months, gathered struggelers, and was closed as stale without a line of code. And whether software echo cancellation, the Ava approach, becomes standard in the voice stack — because if it does, half of part two of this episode evaporates.



Herman

The other thing to watch is the direction it all resolves from. As DIY speech-to-text gets better, the wrong satellite winning becomes merely annoying instead of fatal — same as Google today. The arbitration problem doesn't go away. It just stops being visible, which is how the closed ecosystems have been living with it for a decade. Our thanks, as ever, to Hilbert Flumingtop for producing the show. If this was your kind of episode, go back for episode fifteen, AI Gets Personal; and episode two, Local STT For AMD GPU Owners. This has been My Weird Prompts — no question too obscure, no rabbit hole too deep. If you enjoyed this one, leave us a review wherever you listen; it helps.



Corn

For Daniel's friend: one Voice PE in the kitchen covers ninety percent of households. The question isn't how to match three Echos — it's which of those three rooms you'd actually miss.



Herman

See you soon.

SOURCES

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

1. **GitHub, home-assistant/core issue Duplicate wake-up detection not optimal, opened 2025-12-02, closed 2026-08-24; retrieved 2026-11-13. [primary]**
   https://github.com/home-assistant/core/issues/157762

2. **GitHub, esphome/home-assistant-voice-pe issue Voice PE triggers itself while speaking, opened 2026-01-30, updated 2026-10-27, state open; retrieved 2026-11-13. [primary]**
   https://github.com/esphome/home-assistant-voice-pe/issues/537

3. **GitHub repo, dstrout/wyoming-echo, commit 7073e78, MIT licence; retrieved 2026-11-13. [primary]**
   https://github.com/dstrout/wyoming-echo

4. **Home Assistant official docs, Set up a fully local voice assistant; retrieved 2026-11-13. [primary]**
   https://www.home-assistant.io/voice_control/voice_remote_local_assistant/

5. **and related `assist_satellite` trigger/condition docs (listening, processing, responding, idle, is_listening, is_processing, is_responding) [primary]**
   Home Assistant docs repo, current branch, docs snapshot updated 2026-11-02; retrieved 2026-11-13.
   https://github.com/home-assistant/home-assistant.io/blob/current/source/_triggers/assist_satellite.responding.markdown

6. **Home Assistant docs, ESP32-S3-BOX-3 wake word customization (`micro_wake_word_model` substitutions); retrieved 2026-11-13. [primary]**
   https://github.com/home-assistant/home-assistant.io/blob/current/source/voice_control/s3-box-customize.markdown

7. **GitHub release notes, Ava 0.4.8, software acoustic echo cancellation; retrieved 2026-11-13. [primary]**
https://github.com/knoop7/Ava/releases/tag/0.4.8

8. **Home Assistant Community, thread started 2026-08-03 by Linwood Ferguson; retrieved 2026-11-13.**
https://community.home-assistant.io/t/vpe-multiple-satellites-which-wins-on-wake-word-does-not-work-well/1019767

9. **Big Iron, Multi-Room Voice and Audio Design: Music Assistant Satellites Without a Closed Ecosystem, pricing dated August 2026; retrieved 2026-11-13.**
https://www.bigiron.cc/guides/multi-room-voice-and-audio-design-music-assistant-satellites

10. **EchoMuse docs, Stage 3**
Echo cancellation (AEC); retrieved 2026-11-13.
https://github.com/wilbowes/EchoMuse/blob/main/docs/voice-pipeline.md

11. **Home Assistant Community, thread started 2024-07-13 by super-qua, Multiple s3-box devices in different rooms and the wrong one triggers; retrieved 2026-11-13.**
https://community.home-assistant.io/t/multiple-s3-box-devices-in-different-rooms-and-the-wrong-one-triggers/749237