Episode Book

EPISODE #6486

Why Voice Satellites Need Integrated Audio

PUBLISHED	7 November 2026 at 18:19 UTC
DURATION	24:20
LISTEN	myweirdprompts.com/episode/home-assistant-voice-satellites-92231/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.26
BOOK GENERATED	8 November 2026 at 07:32 UTC

SYNOPSIS

Building a private voice assistant for hands-free home control hinges on acoustic hardware design. This look at Home Assistant voice satellites explores why separating microphones from external speakers breaks echo cancellation and follow-up queries. From the official Voice Preview Edition and custom microWakeWord training to DIY reSpeaker XVF3800 setups, the software trade-offs of Linux Voice Assistant show why hardware subtraction matters far more than just picking a nice speaker.

TRANSCRIPT



Corn

Here's a number: sixty-nine dollars. That's what the official Home Assistant voice box costs, and it still has the word Preview in its name. Daniel wants to know if that's the one to buy, or if he should build something better — and he's coming to this with a home advantage. He's already trained a custom wake word. "Hey Yossi." Both formats, microWakeWord and openWakeWord, sitting there ready.



Herman

He did the homework before the shopping, which is the right order. Most people buy the box first and discover the wake word situation later, which is like buying a lock and then deciding what key you'd like.



Corn

The setup, for anyone joining us: Daniel and Hannah are parenting Ezra, hands full roughly always, and a voice assistant that works from across the room with a child on one arm is the actual goal. The software side he's already mapped — the old Wyoming satellite project got archived in January, replaced by Linux Voice Assistant talking the ESPHome protocol, while Wyoming itself still carries the speech services underneath. His questions are pure hardware. Integrated mic-and-speaker, or separated out? And which products, at what price, given what Home Assistant supports today.



Herman

And we should say up front, prices are approximate — we checked them on the seventh of November, and this stuff moves monthly.



Corn

So let's get clear on what a voice satellite actually is in Home Assistant terms, because it's not a smart speaker and it's not the general far-field microphone problem either.



Herman

A satellite is a thin client. A device in a room with a microphone and a speaker. It listens for a wake word, streams audio to Home Assistant, Home Assistant runs the speech-to-text, decides what to do, sends back a spoken reply. The intelligence lives in



Herman (continued)

the server; the satellite is ears and a mouth.



Corn

Which is why I bristle a little when this gets lumped in with far-field mic engineering. The classic problem is capturing clean speech from anywhere in a noisy room. A satellite has a narrower job: detect one specific phrase, then pipe audio. Capture still matters, but the problem unique to a satellite is the device hearing itself.



Herman

Echo cancellation, yes. While the speaker is playing a reply or music, the microphones are still live, and without something subtracting the playback, the device transcribes its own voice and tries to answer itself. That's the technical heart of the whole integrated-versus-separated question, and we'll get there. But the software framing matters first, because it constrains the hardware. Three routes. An ESP32 device on ESPHome firmware, like Home Assistant's own box, which runs microWakeWord on the device. Linux Voice Assistant, the new thing — experimental, runs on a Pi or similar, talks to Home Assistant over the ESPHome API, and can run your wake word locally in either format. Or the streaming approach, where a device sends audio to Home Assistant and openWakeWord runs on the server.



Corn

And a correction to our own back catalogue. The Pi build we walked through in April leaned on that Wyoming satellite program, so don't follow it now. But Wyoming is not dead. The satellite program retired. The protocol still carries Whisper, Piper, openWakeWord, Speech-to-Phrase. It's the plumbing, and the plumbing stayed.



Herman

The satellite above it got replaced. Right. So for Daniel specifically: Linux Voice Assistant is the one route where either of his Hey Yossi files is usable, because it runs both formats on the satellite itself. You put a tflite model in a wake word directory with a small JSON file naming the type, and that's the shape of it — though we couldn't find a step-by-step procedure for custom models. On an ESP32 box it's narrower. Only the microWakeWord file applies, and the Home Assistant docs say a custom openWakeWord model won't show up in an ESPHome device's wake word selector. And another correction: we once said a custom wake word takes fifty to a hundred samples. We can't support that. Guides talk about thousands of generated clips.



Corn

Which suits him fine. Let's dig into the hardware question, then, and the echo cancellation piece, because that decides almost everything.



Herman

Here's the mechanism. Hardware echo cancellation works by subtraction: the chip takes the signal it's currently playing out the speaker, uses it as a reference, and subtracts that from what the microphones hear. What's left is, ideally, you. The critical word in that sentence is reference. The canceller has to know what the speaker is playing.



Corn

And the reference only exists if the speaker and the microphone share a box.



Herman

Or at least a signal path. The reSpeaker XVF3800 project — that's the XMOS board people are building satellites on — says it in one line of the README: an external speaker like a Sonos, quote, "gives no echo signal to the XVF3800 canceller." So the moment you separate, you must turn off follow-up questions, or the device can hear its own response as a new command, and the stop word timing drifts.



Corn

That's a brutal trade for a parenting use case. Follow-up questions are exactly what you want with your hands full. "Hey Yossi, what's the weather." "Also tomorrow." If the second sentence re-triggers the device because nobody could subtract its own reply...



Herman

You get the assistant answering itself, possibly mid-answer. It's the classic feedback loop, just in software instead of a microphone in front of a PA stack.



Corn

Which is why I'd come down early and say: integrated. For Daniel, integrated. The separated route buys you a nice speaker, and costs you the one feature that makes a hands-free assistant hands-free. In principle, a better speaker wired into the device's own audio output keeps the reference, because the chip still knows what it's sending. We didn't find anyone testing that, but it's the sensible way to get more volume.



Herman

I'm with you, with a caveat that'll take a minute. The caveat is that integrated is necessary but not sufficient, because the echo cancellation on these devices is not a solved quantity. Look at the Voice Preview Edition's spec sheet and it's impressive on paper: an XMOS XU316 chip handling echo cancellation, stationary noise removal, auto gain control, dual microphones, a line-out. Then look at what owners report. There's a forum thread that ran from January to September this year, starting with a developer saying the XU316 wasn't cancelling speaker audio on custom firmware. Someone reproduced it on a Satellite1 that transcribed its own speech. No resolution in the thread. Meanwhile a reviewer running six units says the echo cancellation and noise reduction are decent.



Corn

Six happy units against a nine-month thread of the opposite. That's not a spec dispute, that's a firmware lottery.



Herman

I wouldn't go that far. That thread is largely about custom firmware, so it doesn't settle how a stock unit behaves. Treat it as unsettled. We found no systematic test.

L

Larry

Somewhere, across a quiet room, a wake word waits. Hey Yossi. Whispered. Unhurried. Almost a secret. And now, the answer to that secret: Echo Hush. The luxury of being heard. Echo Hush listens for your every word, then gently, luxuriously, subtracts it from itself, so you never have to say it twice. Pairs beautifully with any speaker you love, or none at all, and leaves you with nothing but silence. Pure. Echoed. Hush.

L

Larry

Echo Hush. For the room that hears too much. Echo Hush.



Herman

That's — no. That ad has it exactly backwards, and it's bugging me. It says the device "subtracts it from itself," which is garbled, but worse, "pairs beautifully with any speaker you love, or none at all" is precisely the configuration where echo cancellation cannot work. If the audio comes from a speaker the device doesn't control, the canceller has no reference signal. It subtracts nothing. You'd have to disable follow-up questions, which for a hands-full parent is the whole feature. There is no luxury product that fixes physics.



Corn

Let it go, Herman. It's a whisper commercial.



Herman

It whispered wrong.



Corn

The design workaround here is interesting, though. Linux Voice Assistant doesn't actually trust echo cancellation in the general case — it sidesteps it. The behaviour doc says the wake word is ignored while a reply is speaking. The stop word works during cancellable audio. Music ducks to fifty percent rather than getting fully cancelled.



Herman

There's no general barge-in. You can't interrupt it mid-sentence the way you can with an Alexa, beyond the stop word. The effect is that if the device isn't listening while it speaks, it doesn't strictly need to cancel its own voice. It trades away interruptibility for robustness. And for the lullaby-at-nap-time scenario, note that forum users report wake words failing during music on some devices, with muting the media player during detection as the workaround.



Corn

Which is very much my kind of engineering. Don't solve the hard problem; arrange never to meet it.



Herman

You'd have gotten along fine with the protocol designers.



Corn

There's one more integrated route nobody's validated. USB speakerphones. The Jabra Speak range — the 510, the Speak2 — these are conference call pucks. Microphone and speaker in one box, vendor-claimed onboard echo cancellation, USB. In principle the reference signal stays inside the device, same as the XMOS boards.



Herman

In principle. We found nobody documenting one running with Linux Voice Assistant. There's one older forum report of a Jabra Speak 410 working with Rhasspy years back, where TV noise was the complaint. So it's a someone-should-try-this, not a recommendation. Though I'll say this for it — conference hardware is engineered for exactly Daniel's scenario: someone across a table, talking at conversational volume, while the puck itself is talking.



Corn

The one room a speakerphone is guaranteed for is the room it was designed for. What's the price on that route?



Herman

A Speak 510 runs about a hundred thirty-four to a hundred seventy-two dollars depending on retailer, the Speak2 55 closer to two hundred to two-forty, and either way you still need a Pi or mini PC to plug it into. At that money you're well past the purpose-built options.



Corn

So it's an idea, not a purchase. Let's get to the purchases, then. If echo cancellation is the deciding factor and integrated wins, what does the landscape actually look like today?



Herman

Start with the reference device, because everything else is measured against it. Home Assistant Voice Preview Edition. Sixty-nine dollars, or fifty-nine euros. Dual microphones, the XU316 doing echo cancellation and gain control, a 3.5 millimetre line-out. The built-in speaker is for voice replies; the music job goes to that jack. Still named Preview Edition, and we found no successor. The catch for Daniel: it ships with three wake words, and getting Hey Yossi on it means taking over the firmware in ESPHome and pointing it at his microWakeWord model. Home Assistant's page calls making custom words not realistic for most users. He's already done the hard half, but forum users report that takeover breaking across ESPHome versions.



Herman

The forum grumbles are worth knowing about but not over-weighting: one owner describes needing to be within about a metre and a half, with a raised voice. A report, not a measurement. We found no independent far-field range test on any of these.



Corn

A metre and a half with a raised voice is roughly the interaction radius of shouting at your phone.



Herman

Which is cheaper and already in your pocket. So keep it as a caution, not a verdict. Next, the DIY escalation: Seeed's reSpeaker XVF3800 with the XIAO ESP32-S3. Four microphones instead of two, roughly sixty-two to sixty-nine euros at Openelab. Seeed claims capture up to five metres, which is the manufacturer's number. This is a board rather than a finished speaker, so budget for a small speaker on its own output and a case if yours doesn't include one. The catch is the software. ESPHome support is a community package — explicitly marked under development and at your own risk, and it needs ESPHome 2026.10 or later. Seeed's own wiki points to it, but it's not official.



Corn

"At your own risk" in a README is the open-source way of saying we love you, goodbye.



Herman

It's honest, at least. Whether four microphones actually beat two is another untested claim, by the way. No independent comparison exists. Four mics give the beamforming more to work with in theory, and theory is all we've got.



Corn

Middle option?



Herman

The Seeed ReSpeaker Lite. Two microphones, same XU316 chip, around thirty-four dollars for the kit in a 2024 review — current price I can't confirm. The Linux Voice Assistant README actually recommends it, paired with at least a Raspberry Pi 3. With a note that it currently has a problem on the Zero 2W, the small Pi. Below that there's the reSpeaker 2-Mics Pi HAT, the budget floor, and it uses a much more basic audio codec. We recommended that HAT back in April, and for this job it's now the floor, not the foundation.



Corn

And the premium end?



Herman

FutureProofHomes Satellite1.1, assembled. A hundred thirty-two ninety-nine — and shown as sold out as of today. XMOS-based, and the Linux Voice Assistant README recommends its HAT board on a Pi Zero 2W. The pricing history is murky, though — an older forum comparison quoted about seventy dollars for an earlier configuration, so the figures conflict and I wouldn't build a plan around either number.



Corn

The Atom Echo, for the record, is the learn-ESPHome toy. Single microphone, a tiny speaker, no DSP, seventeen or eighteen dollars. We covered it back when we did the per-room satellite episode — fine for finding out whether you enjoy this hobby, not a far-field device.



Herman

So the shortlist, roughly: Voice PE at sixty-nine as the default, if he'll take over its firmware for Hey Yossi. ReSpeaker Lite, around thirty-four on that older price, plus a Pi if he wants Linux Voice Assistant and either of his two model files. XVF3800 if he wants four mics and accepts community firmware, Satellite1.1 if he can find one, Atom Echo only as a toe in the water.



Corn

That's five. Good number for a shortlist. Now the part that matters more for Daniel than for anyone else buying one of these: what happens with more than one.



Herman

This is where I have to correct us again, because we described this wrongly on air. There is no satellite grouping feature. There's no nearest-device arbitration. What Home Assistant actually has is a five-second cooldown on the same wake word: when two devices hear "Hey Yossi," the second one gets a duplicate wake-up result and the first device to reach speech-to-text wins. Not the closest one. The fastest one.



Corn

So if he puts a satellite in the kitchen and one in the hallway, the answer might come from the wrong room.



Herman

Yes, whichever gets there first. In a small apartment, the devices will be close enough together that both hear him regularly. So place them to overlap as little as possible, and expect first-past-the-post, not smart.



Corn

And no built-in speaker recognition, I assume — so it can't tell Daniel from Hannah.



Herman

Nothing built in that we could find, and that corrects us again — we said Home Assistant could do user-specific voice recognition. Third-party projects exist, but nothing official. Two wake words and two pipelines per satellite have been supported since the October 2025 release, though — so you could, for instance, run one local pipeline and one cloud pipeline and switch by wake word. That's the closest thing to a per-person or per-context split today.



Corn

There's also the zero-hardware option, which nobody ever considers because it's boring. The Assist Satellite app for Home Assistant OS runs the same runtime on the Home Assistant host itself.



Herman

If the box is in the room. You have to pass a microphone and speaker through to the machine, so it suits a host sitting in the living room, not one in a cupboard. If his box doesn't live where the family lives, Linux Voice Assistant on any spare Pi does the same job. Either way, he could test the stack with a USB microphone he already owns before spending anything.



Corn

That's the first thing I'd do. Find out the software works with Hey Yossi at all, using hardware that's already in a drawer, before ordering a sixty-nine-dollar preview of anything.



Herman

Practical to the last, Corn.



Hilbert

Jabra Speak 510. A hundred and forty-nine dollars, 2014, from an office supplier on Ibn Saud Street, and I have never gotten along with it. Herman's four-microphone talk earlier — you can keep it. I had a fish tank in the living room, a big one, four feet, and glass reflects sound like a billiard table. Every echo canceller I put in that room gave up. The Jabra went back. So did a Philips. So did a Logitech, though the Logitech was mostly the tank as well. Three returns, one season.



Corn

What finally worked?



Hilbert

A car boot unit. No name on it, five shekels, and I mounted it upside down under a shelf, because the bracket only bent that way. Corn, that's the part you'll like: it worked. Nobody knows why. I ran it for six years without touching it, and I resented every day of not having a failure to diagnose. The man who sold it to me said his uncle used it in a butcher's shop, and that the uncle could hear a whispered order from the cold room door. Nine metres, he claimed, and he'd measured it. Once.



Herman

Upside down changes the microphone geometry relative to the surface it's near — the shelf becomes part of the acoustic design whether the manufacturer intended it or not. Ceiling and wall proximity shifts what the canceller sees. It's not crazy that orientation fixed it.



Hilbert

Possibly. I still have the tank. New devices go next to it now, before anything else — if a unit transcribes its own output with the filter running behind it, it goes back on the shelf. I call it the aquarium test. It has never once been wrong about a device, only about me, because I keep buying them anyway. The claws don't help with the tiny buttons, but that's a separate grievance. The 510 itself I sold at the car boot, to the same man. He didn't want it either.



Corn

So your position is that integrated is right, but the room has a vote.



Hilbert

And nobody counts the ballots. The XU316 sitting against a window, or two inches from a wall — the canceller is tuned for its own enclosure, and Daniel's apartment will do what it likes with the leftovers.



Corn

Put the thing on a table away from glass, is what I'm hearing. Free advice, no car boot required.



Hilbert

The five shekel solution is still the best one, but yes, the table.



Corn

The aquarium test — that's the missing product review. Every voice satellite shipped this decade, scored by whether it can survive standing next to a fish tank.



Herman

It's honestly a more rigorous protocol than any of the forum threads we discussed.



Corn

So where does this leave Daniel. The hardware landscape is unsettled in a way the marketing won't admit — the flagship is still called Preview Edition, the four-mic board runs on community firmware marked at your own risk, the premium option is sold out, and the one measurement we couldn't find is the only one that matters: does it hear you from across the room. My guess is he tests Hey Yossi on hardware he already owns, then buys one Voice PE, keeps the receipt, and finds out within a week whether it hears him across the room with the washing machine on.



Herman

And the software will move faster than the hardware either way. Linux Voice Assistant is experimental and improving; the XVF3800 package is under active development. The one thing Daniel has that almost nobody does is the wake word, trained and ready in both formats. Which means the box he picks only has to run it.



Corn

Forward-looking thought, then: the interesting question isn't which box to buy this month. It's whether anyone ever builds the arbitration layer — some way for the kitchen satellite to notice it's the one being spoken to. First-past-the-post is fine in a small apartment with one wake word. It's not fine for what this ecosystem wants to become.



Herman

And someone should run the Jabra under Linux Voice Assistant. There's a five-shekel answer waiting somewhere.



Corn

Producer credit where it's due — Hilbert Flumingtop has been at the desk, as ever. If you've built a voice satellite, especially with a custom wake word, leave us a review and tell us what hardware you used and whether the echo cancellation actually held up. This has been My Weird Prompts — no question too obscure, no rabbit hole too deep. Find the show at my weird prompts dot com. We'll be back soon.



Herman

See you tomorrow, Corn.



Corn

Tomorrow.