Episode Book

EPISODE #6547

USB Far-Field Mics and the Echo Cancellation Catch

PUBLISHED	13 November 2026 at 12:19 UTC
DURATION	21:07
LISTEN	myweirdprompts.com/episode/usb-far-field-mic-arrays-36994/
HOSTS	Corn, Herman, Orson, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.26
BOOK GENERATED	14 November 2026 at 06:47 UTC

SYNOPSIS

A far-field microphone array is really a small computer with microphones attached, and the DSP chip behind them is the actual product. For a DIY Home Assistant and Whisper satellite, USB Audio Class boards like Seeed's ReSpeaker XVF3800, the older ReSpeaker v2.0, and the ReSpeaker Lite offer onboard beamforming and noise suppression, unlike generic AliExpress arrays or the PlayStation 3 Eye. The catch is acoustic echo cancellation: the chip can only cancel audio it plays through its own output, meaning separate speakers break barge-in.

TRANSCRIPT



Corn

Daniel's new prompt is about the part of a voice assistant nobody demos. He's wiring up his own satellite, a Pi or a small PC running Home Assistant with a wake-word engine feeding Whisper, and he wants to know which standalone far-field microphones actually perform. Beamforming, some noise rejection, echo cancellation if he can get it, and over USB, not Bluetooth. He's also noticed what everyone who browses AliExpress notices, quote, quality varies a lot. He wants the ones validated by people running them, not the ones with nice product renders. And his benchmark is the one everyone's friend has: the Amazon Echo, the thing that hears you from across the room over its own music. His hunch is that there's nothing magical about its hardware.



Herman

His hunch is right, and we'll come back to why. The original Echo had a ring of seven microphones, but microphones are the cheap part. There's a second half to his question, and it's the better half. Does he buy one device that is both microphone and speaker, or pair a far-field mic with a separate speaker, and what are people actually pairing? That sounds like a shopping question. It's an architecture question.



Corn

They're the same question. Which mic, and what it's plugged into. The answer to the second one decides what the first one can even do.



Herman

Then the framing matters, so let me set it properly. A far-field mic array is not really a microphone. It's a small computer with microphones attached. The mics themselves are the cheap part. The chip behind them is the product. Beamforming, noise suppression, automatic gain, echo cancellation, all of it runs on the device, on the silicon, before Linux ever sees a sample.



Corn

A microphone with firmware. My favorite genre of microphone.



Herman

And that's why USB matters here. These arrays show up as USB Audio Class sound cards. Driverless on Linux, no ceremony, and that's exactly what Home Assistant and a Whisper pipeline expect. A sound card. Bluetooth is the trap door. Codec latency, pairing behavior, buffering. Wake-word engines and barge-in do not survive a link that holds samples in a queue.



Corn

Barge-in, meaning talking over it. The thing every commercial assistant does and every DIY one fails at.



Herman

Two chip families basically own this class. The XMOS XVF3800 is the current generation, and the older XVF-3000 and XU316 silicon still powers plenty of shipping products. Same chips inside very different boards at very different prices. The one thing to hold onto going in: whatever the microphone can do, echo cancellation is only half a feature. It depends entirely on what the mic can hear the speaker playing, and that is Daniel's second question hiding inside his first.



Corn

Then the microphone first, because that's where the spec sheets mislead most.



Herman

The current leader is Seeed's ReSpeaker XVF3800 USB four-mic array. Launched July twenty twenty-five. The XMOS chip does the full stack. Beamforming, de-reverberation, direction of arrival, neural noise suppression, sixty decibels of automatic gain, and genuine acoustic echo cancellation. Four mics, pickup out to about five metres, and a maximum sample rate of sixteen kilohertz.



Corn

Sixteen kilohertz sounds low until you remember what it's for.



Herman

Speech recognition doesn't want more. Whisper downsamples anyway. You're not losing anything the model would use. What you're buying is the chip deciding, at the source, which sounds are you. Launch pricing was fifty dollars for the bare USB board, about fifty-four cased, and roughly fifty-six on AliExpress at launch. By this autumn people were quoting sixty-five before VAT in forum comparisons, and Amazon listings at eighty cased in the spring.



Corn

The case appreciates faster than the microphone.



Herman

Scarce, too. Several UK and EU shops had it sold out or backordered. Now here's the part that catches people, and it caught me with my jaw slightly open on the wiki. There are two versions. The plain board ships with USB firmware. The version with a XIAO ESP32-S3 attached ships with I2S firmware, meaning you plug it into your Pi and it does not appear as a USB device at all. It isn't broken. It's just speaking a different protocol out of the box.



Corn

So someone buys the more expensive version, plugs it in, sees nothing, and concludes they've been sold a dead board.



Herman

That exact person exists in every thread about this device. The fix is a firmware flash. You hold the Mute button while powering it up, then flash over with the DFU utility. There are images for it. The standard one, a sixteen-kilohertz six-channel build that exposes the raw mic channels...



Corn

...some of which need enabling in a mixer, because nothing in Linux audio can ever be simple.



Herman

Some of them, yes. And a forty-eight-kilohertz two-channel image if you want plain stereo. It's USB Audio Class underneath, so Linux takes it with no driver. But you need a udev rule for the control tool, and the speaker output is famously quiet by default. File that last one away, it becomes load-bearing later.



Corn

Filed under speaker problems, subcategory quiet.



Herman

Real-world validation, because that's specifically what Daniel asked for. The best data point going is a Home Assistant forum comparison from this autumn. The XVF3800 beat the Voice Preview Edition at long range in quiet rooms, and the reviewer's phrase was almost too sensitive.



Corn

Almost too sensitive. Meaning what, precisely?



Herman

Meaning two to four of their seven satellites woke at once. You say the wake word in the kitchen and the hallway and the living room all raise their hands at the same time. That's a cross-room coordination problem, and it's a separate conversation from today's, so I'll leave it there. The other finding I find more honest: beamforming helped with an off-angle noise source, a fan off to one side. It did nothing when the person stood between the noise and the device.



Corn

Because the beam steers toward the talker. The noise comes along for the ride.



Herman

That's the honest limit of the whole feature. Beamforming gives the device a sharper ear pointed at you. It doesn't teleport the dishwasher away.



Corn

You've got the face you get when a spec sheet says the word neural, by the way. Just so you know.



Herman

I do like a marketing department that finally means it. Sixty decibels of automatic gain means the thing can hear you whisper from across the room, whether you wanted that or not. Very attentive roommate. The cheaper tiers now, because not everyone wants fifty dollars of microphone. The ReSpeaker USB Mic Array v2.0. Older XVF-3000 chip, four mics, about five metres, echo cancellation, plug and play on Linux, macOS and Windows. Retail pricing is inconsistent. Farnell lists it at fifty-nine pounds including VAT, which is more than the newer, better board costs. The supply chain hasn't noticed the successor exists.



Corn

The old stock charging a premium for being old.



Herman

Then the ReSpeaker Lite. XU316 chip, two mics, USB Audio Class, the same DFU dance if you want USB firmware instead of I2S, and you want firmware newer than 2.0.5. What there isn't is Pi test data. Almost nobody has published any. Two mics also caps what beamforming can do for you geometrically.



Corn

Beamforming with two microphones is a compass with one needle. You know something is out there.



Herman

The Pi HATs, the four-mic and six-mic boards that sit on the GPIO header, are plain microphone boards. No onboard DSP at all. Every sample goes to the Pi, the same Pi running the wake-word engine and handing off to Whisper. And the Matrix Voice, which people still ask about, is listed as no longer manufactured and was a HAT-style board anyway, not USB.



Corn

Which leads us to AliExpress, which is Daniel's actual question. He's seen the variance himself.



Herman

It's structural, not seller roulette. AliExpress carries the XVF3800 itself, the miniDSP UMA-16, Sipeed arrays, and a whole wall of generic four, six, eight, sixteen-mic USB arrays. Two separate issues there. Whether the XVF3800 listings on AliExpress are genuine Seeed units is unverified. I haven't seen anyone confirm it either way, and I'd want the Seeed branding checked closely before trusting a discount. The generic arrays are the bigger trap. The ones with the impressive mic counts often have no onboard DSP whatsoever. Sixteen microphones with no chip behind them is sixteen ways to hear the same noise.



Corn

The mic count is the spec-sheet number. The DSP is the product.



Herman

Now the classic hack, for the record, because it's the ghost at every one of these parties. The PlayStation 3 Eye camera. Four-mic USB array, sold used for seven to thirteen dollars, works on Linux with some ALSA configuration. No DSP, no echo cancellation, fixed gain. Every benchmark praising it dates from twenty seventeen to twenty nineteen, before this entire generation of XMOS chips existed.



Corn

A ten-dollar camera with four ears. It deserved its moment and it had it.



Herman

Hold on. Before we leave the hardware behind, we've got a correspondent out there. And now let's go live to our correspondent, who has been standing by at the scene.

o

Orson

Good evening. I stand before you in the shadow of the microphone array, where the far field is a tragedy in five acts, and the chorus is a rasp of rolling Rs. The chip is the computer, the computer is the chip, and the reference signal is a ghost that haunts the echo cancellation.



Corn

Orson, which array are you standing in front of?

o

Orson

Beamforming, de-reverberation, automatic gain of sixty decibels, all of it whispers in the dark. The USB cable is the only honest messenger, driverless and true. Here ends the report. Orson, in the far field.



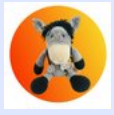
Herman

Where else.



Corn

Reference signal. Orson buried the lede, as usual, but he buried it in the right place.



Herman

He did, actually. Here's the mechanism, and it reframes Daniel's entire speaker question. Echo cancellation is not a property of the microphone. The canceller needs a far-end reference, meaning the exact audio being played out the speaker, fed in on the chip's input. The XMOS user guide is explicit about this. Two consequences follow. The canceller takes a few seconds to converge when playback starts, so the first moments of barge-in are the worst ones. And any external gain on that reference path has to be matched in a



Herman (continued)

parameter, or the canceller misjudges how loud its own voice is.



Corn

It has to know what it said in order to know what to ignore.



Herman

That's the entire trick. And in USB mode, the chip's own output is what drives the speaker. Follow it to the end and you get the part nobody puts on the box. The array can only cancel audio it plays itself. Through its own sound card, its own jack. If the music or the TTS replies come from another sound card, or from a separate smart speaker on the shelf, the canceller never hears them. It isn't a degraded array at that point. It's just an array.



Corn

And you lose barge-in.



Herman

Completely. The wake-word engine is listening to your music at full volume with nothing subtracted. A Home Assistant forum user put it flat. None of the Voice Preview Edition, the FutureProofHomes Satellite1, or the XVF3800 gets an echo reference when an external speaker is used. None of them. Even the purpose-built hardware hits this wall.



Corn

So when Daniel asks one device versus mic plus speaker, the real question underneath is whether the canceller ever receives the reference.



Herman

And the uncomfortable part is that whether USB playback reliably serves as that reference under Linux isn't documented by Seeed. The architecture says it should work, playback goes out through the chip, the chip hears its own output. But nobody has published a careful confirmation, and I've learned not to assume a datasheet's implications are free.



Corn

That's a big hole under a load-bearing wall.



Herman

It is. And notice how the quiet speaker output connects. If you boost that output externally, you've just added gain to the reference path, and the guide says that gain has to be matched in a parameter. The quiet default isn't just an annoyance. It's part of the calibration.



Corn

Now the people who solved this by design. Speakerphones.



Herman

Conference speakerphones. Jabra Speak 750, full duplex, echo cancellation, three sixty pickup. Anker PowerConf S330, four mics, three sixty pickup, auto gain, and notably no explicit echo-cancellation claim I could find for the Anker. But the geometry is the point. The speaker lives in the same box as the mics, so the echo canceller hears exactly what's played. Reference solved by construction.



Corn

Built for people around a table. Which is both the pitch and the problem.



Herman

Users do run the Anker units and the Jabras as Home Assistant satellites on Pi 3s, Pi 4s, mini PCs. The recurring complaint is weak wake-word pickup. These things are tuned for someone two metres away who accepted a calendar invite. Not someone calling from the kitchen with the extractor fan on. The failure reports pile up too. A Jabra 510 on a Home Assistant Yellow gave no audio and kept disconnecting, and the older 410 and 510 needed manual ALSA setup under Rhasspy back in that era.



Corn

Here's the thing I keep circling back to in all of it. Nobody has tested wake-word reliability during music playback on any of these speakerphones. Not one source anywhere.



Herman

Not one. And that's the single scenario that decides whether barge-in works in a real living room. Everything else is desk testing with the room quiet. Corn, honest question, and I mean it. Would you run a speakerphone, or the array driving its own speaker?



Corn

If the canceller needs the reference, I want the reference guaranteed by wiring, not by hope. The speakerphone tempts me, but I don't want a meeting-room brain in my hallway thanking me for my patience.



Herman

I'd have bet on you defending the separate speaker, honestly. You like the old repairable option with parts you can replace.



Corn

I like the old option when it works. A separate speaker with no reference is just a nicer way to deafen the wake word.



Herman

Then there's the third path, and people do take it. Software cancellation on the Pi. PipeWire has an echo-cancel module, and there's the WebRTC canceller. When the hardware gets no reference, you subtract the echo in software after the fact. It works, mostly. But it's a fallback, not a free lunch. You're doing the DSP on the same box that's running everything else, and it's never as clean as the chip doing it at the source, before the signal ever leaves the device.



Corn

Software doing the job the hardware was designed for, on a computer that already had a job.



Herman

The ground is moving under all of it anyway. The old Wyoming satellite project was archived at the end of January this year. The Open Home Foundation's Linux Voice Assistant is where the momentum is, ESPHome API underneath, still experimental, and by user reports still rough on a Pi 5. Community forks are adding XVF3800 USB support. I installed one of those forks the week it was updated, naturally, and naturally my satellite stopped announcing anything until I re-flashed it.



Corn

I remember the silence. You were very calm about it for someone who broke the hallway.



Herman

The damage report is part of the enthusiasm. The comparison worth holding onto is the Voice Preview Edition itself. Sixty dollars before VAT, XMOS XU316 inside, but it's an ESP32 network satellite, not a USB mic, and there's a forum thread reporting its echo cancellation not working. Even the integrated product with the right chip from the right company doesn't get this for free.



Corn

A useful humility marker. Right chip, right product, still haunted.



Herman

So the popular pairings that actually exist out there. The XVF3800 driving its own speaker through its own output. A conference speakerphone as the all-in-one, with the wake-range caveat. Or far-field mic plus separate speaker with software cancellation, which is where barge-in goes to die. You pick by which failure you can live with. Corn, does anyone actually still run the PS3 Eye? Seriously. Is there anyone left?



Corn

Given what those forum threads looked like, I'd guess a few holdouts, but I've never met one.



Hilbert

You have. Me. One of them, in the drawer, with both foams. The grey original and the black replacements that came later. The grey is the one that matters. The black ones sit on the high frequencies like a duvet. I ran it on a Pi for wake words two winters ago, and it was fine until the radio was on. Then it heard the radio say everything. I'd say "lights" and it would wait politely for the newsreader to finish.



Corn

No echo cancellation, so the radio wins.



Hilbert

The radio always won. I kept it because it was nine pounds and it has a camera I've never once used. It's in the drawer next to the grey foams, which I don't throw away, because you can't get the grey ones any more.



Herman

That's the PS3 Eye in one story. Four good ears and nothing behind them deciding what to ignore.



Corn

Which brings us back to Daniel's Echo. What's actually the trick?



Herman

The trick is that the Echo is the speakerphone case done properly. It's playing its own music through its own speaker, so its echo canceller always has the reference. It never has to guess what it said. Add a good array and years of tuning on the processing, and that's the magic. Daniel's right that the hardware isn't exotic. What's hard to copy is that



Herman (continued)

the whole loop lives in one box and was tuned as one box.



Corn

So, if Daniel's building one satellite this month?



Herman

The XVF3800 on USB firmware, with the speaker plugged into the array's own output, not into the Pi's. Raise the volume in software with the reference gain in mind. Then do the test nobody has published: play music at a normal level and see whether it still hears the wake word. If it fails that, a conference speakerphone is the fallback, and you accept a shorter wake range.



Corn

And whatever he buys, test it with the music on.



Herman

That's the whole episode in one test.



Corn

Our thanks, as ever, to Hilbert Flumingtop for producing the episode, and for the foams.



Herman

If this one was useful, a review wherever you listen helps other people find the show. This has been My Weird Prompts. No question too obscure, no rabbit hole too deep.



Corn

See you next time.