Episode Book

EPISODE #6724

Controlling a Web App While Walking

PUBLISHED	21 November 2026 at 14:00 UTC
DURATION	25:09
LISTEN	myweirdprompts.com/episode/controlling-web-app-while-walking-26272/
HOSTS	Corn, Herman, Dorothy, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	omnivoice
PIPELINE	V5.36
BOOK GENERATED	21 November 2026 at 19:31 UTC

SYNOPSIS

Controlling a custom web app with a phone tucked away in a pocket runs headfirst into Android's input architecture. While tiny Bluetooth remotes and voice commands seem appealing for speech recording on the go, Chromium drops keyboard events the instant a page loses visibility. This breakdown explores why media session keys succeed where handheld gamepads fail, the drawbacks of voice triggers in speech datasets, and how multi-tap patterns solve mobile ergonomics.

TRANSCRIPT



Corn

Eighty hours of recorded speech. Daniel has been pacing around his house and now the streets of Jerusalem, reading one sentence at a time into a headset. The app he built reads him the sentence, and everything he does with it comes down to about seven actions — record, stop, save and next, retake, skip, pause, play the last take. Seven buttons. The problem is that he physically cannot stand still while doing this, the phone is often in his pocket, and the 8BitDo he bought has been annexed by Ezra, who is the most persuasive negotiator either of us has heard of.



Herman

And the question underneath all of it — what Daniel actually asked — is what would you drive those seven controls with while walking. He's tried or considered the obvious things. The 8BitDo never felt right even before it became a toy. The two-dollar ring remotes from AliExpress feel like party favours. Voice commands have the obvious poison in them: the words "stop" and "retake" can show up inside the sentences he's reading. And then there's the far end of the spectrum — a sensor pad in his shoe he'd press with a toe.



Corn

My kind of question. My only question is whether the shoe thing was a joke.



Herman

Knowing Daniel, it was a quarter-joke. Which is enough to take seriously.



Corn

Ridiculous on the surface, and then you catch yourself thinking about the ergonomics of it at two in the morning.



Corn

He ends with the exact right prompt though — "what would you actually use". Not a survey. A recommendation.



Herman

Then let's actually give him one. But to give him one honestly we have to look at why the obvious answers fail, because the reason they fail isn't the reason he thinks.



Corn

Go on.



Herman

The real constraint here is not buttons. It's the Android input stack. Daniel's app is a web app — a PWA running in Chrome on Android. And Chromium, the engine underneath Chrome, only dispatches gamepad and key input to a page that is visible. The moment that phone goes into his pocket and the screen dims or locks, the page loses focus, and every keystroke and every gamepad button in the world goes nowhere.



Corn

Nowhere as in the browser swallows it, or nowhere as in it lands somewhere else?



Herman

Nowhere useful. The page just doesn't see it. His app does hold a screen wake lock while recording, so the screen shouldn't dim on its own. But one knock of the power button in a pocket, or the app getting backgrounded, and the page is hidden, and a wake lock is released the moment the page is hidden. Visibility and the lock die together.



Corn

So any controller he tries can fail at a layer he can't see. The pad's fine. The browser just stopped listening.



Herman

That's the failure mode to test first, before buying anything. He said the 8BitDo "never felt like quite the right tool" — I'd bet a decent lunch that's partly because it's a two-thumb device being used by one thumb on a moving body. And the moment the page is hidden,



Herman (continued)

the pad is a very well-engineered brick.



Corn

Twenty-five grams of brick. Fifteen dollars on sale.



Herman

Well — it lists at twenty-five dollars, and it does drop to around fifteen when it's discounted, which is most of the time. Seventy-two by forty-one millimetres, sixteen buttons, and it has a proper keyboard mode, so in principle it sends exactly the keys Daniel's app already listens for. But the reviews of that keyboard mode are consistent in one ugly way: the default key maps are inconsistent. Half the job is discovering what the buttons actually send this week.



Corn

A controller that keeps you guessing. Perfect for a man walking.



Herman

There's one exception to the visibility rule, and it's the crack in the whole problem. Media keys. Android routes headset and earbud button presses through its media controls, and a page with an active media session receives those even when the page is in the background. Daniel's app already registers a media session with play, pause and stop handlers, and plays a silent loop to keep it alive — that's why he gets the media notification while recording.



Corn

Wait. So the headset he's already wearing — the one the microphone is on — its button already works while the phone is in his pocket?



Herman

Works right now, today, for one thing: a single press is play-pause, so pause and resume. But Android's mapping goes further: a double press is next track, a triple press is previous track, a long press summons the voice assistant. That's three usable gestures from a button he is literally already touching, because his microphone is on that headset.



Corn

He's been holding a remote this whole time.



Herman

He has. And here's the part where his own app does half the work for free: if he adds next-track and previous-track handlers to the media session — which is a few lines in the app, not new hardware — those double and triple presses suddenly have somewhere to land. Single, double, triple: three actions, zero new hardware, and they survive the pocket. Earbuds with their own tap and swipe controls can add a volume pair on top.



Corn

And mis-taps. Walking doubles everything. You hit a curb, your thumb tells your ear it's a triple.



Herman

That's the honest catch, yes. Double and triple presses are counted by timing windows, and a man striding down a hill with a curious dog is not a laboratory. But the actions he'd map there are the recoverable ones. Retake, skip. A false trigger costs him one take, not the run.



Corn

Which is a good argument for putting "stop the run" nowhere near a double-click. That's the one you want boring and deliberate.



Herman

Put stop on the long hold if anything, or leave it on the phone entirely — the one action you stop walking for. I like that discipline actually. Six things can happen mid-stride, the seventh is the one you're allowed to sit down for.



Corn

It's the ejector-seat principle. You don't want the ejector seat on the same lever as the radio.



Herman

Precisely the principle, yes. And it's worth saying this is how the app is probably structured anyway — record, stop, save are the spine of a session, and retake and skip are the moment-to-moment corrections. The spine should be hard to hit by accident. The corrections can be loose.



Corn

So what about the tools he's actually tried — take the 8BitDo first, since it's earned a proper autopsy.



Herman

The autopsy is short. Two-thumb layout, one hand available, inconsistent keyboard mappings, and the visibility rule killing it in the pocket. Even in its best case — phone out, screen awake — it's a device designed for someone sitting down. The Zero 2, the keychain-sized one, has the same soul in a smaller body.



Corn

And the rings?



Herman

The cheap ring remotes — two to ten dollars on AliExpress, twenty for the branded ones — are almost all generic Bluetooth scroll devices or camera shutters. They send a fixed key or two, often a volume key. So you get one action, maybe two, and then you're into Key Mapper on Android to translate them into something the app understands.



Corn

One action. For a seven-control app.



Herman

You could wear seven rings.



Corn

Herman.



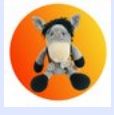
Herman

I'm just noting the option. A man with a ring per verb. "Stop" is the signet.



Corn

He'd lose the retake ring within a week. Then he'd be stuck half-retaking forever. What's Key Mapper, for listeners who don't spend their evenings in it?



Herman

Open source, Android, excellent piece of work. It can use almost any button as a trigger — including a headset button — with long-press, double-press, N-press patterns, and map it to anything. The limits matter though: it doesn't handle gamepad joysticks or triggers. So it's a bridge for keys and audio buttons, not a universal remapper for a gamepad's full anatomy.



Corn

So the rings plus Key Mapper is one action wearing a disguise. Voice commands next, because I actually think this one is subtler than it looks.



Herman

It is subtler, and the subtlety is the poison. Daniel's instinct is right — words like "stop" and "retake" can show up in the sentences themselves, so you can't use plain English commands. The fix people reach for is a nonsense trigger word that never appears in English prompts. And that does work, in one direction: it removes false triggers from the sentences.



Corn

But?



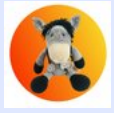
Herman

But the command still lands in the recorded audio. He finishes reading the sentence, says "zorblax" or whatever the trigger is, and every single take now has a nonsense syllable at the end that has to be trimmed off before it goes into the training set. Eighty hours of takes so far. He's talking about a trim operation multiplied across the entire corpus going forward.



Corn

So voice control in a speech corpus app is like printing the editor's notes inside the book.



Herman

That's the shape of it, yes. It's not a reliability problem first — you can get reliability, there are decent on-device wake word engines. It's that the control channel and the data channel are the same channel. It's the only input method on the table that contaminates the thing it's controlling.



Corn

And the timing's awkward too, isn't it? Wasn't Picovoice's free tier—



Herman

Shutting down. Free-tier access keys stopped working end of June this year, no non-commercial tier planned. Which matters because Porcupine, their wake word engine, was the polished option for exactly this. The open alternative is openWakeWord — Apache 2.0, runs on Android, has an automated pipeline for training custom words — but recognition is reportedly hit and miss for some users. So the one input method that dirties the data is also the one with the shakiest tooling this year.



Corn

A bad idea that's also getting harder to execute. Rare to see both failings aligned so neatly.



Herman

There is a workable corner of it — speak the command after stopping, or use silence detection as the boundary. The dataset tools people have built for exactly this purpose, things like Revoxx and Piper Recording Studio, all use keyboard shortcuts and silence detection. Not one of them uses voice control. The people who record speech for a living have collectively voted no.



Corn

That's the strongest evidence in the whole episode and it took one sentence.



Herman

So the real question isn't which gadget. It's what the Android input stack will even let you do while the phone is in your pocket — and we've found one clean answer already, the media session. Now the question is whether anything else survives that constraint.



Corn

Before you get to the shoe, I want the shoe.



Herman

The shoe is coming. But it's bracketed by two categories we haven't cleared: handheld buttons that aren't gamepads, and actual foot hardware. Take handheld first — there's one device in this whole space that I think fits, and it's the Flic 2.



Corn

Daniel owns a Flic.



Herman

He does — he mentioned he has one. And the Flic 2 is doing something structurally right for this problem: on Android it presents itself as a Bluetooth keyboard. HID mode. So the button press arrives at the phone as a keystroke, which means Daniel's app — which already lets you bind any key — just learns it. No bridge software.



Corn

But the visibility rule. A keystroke to a hidden page is dead on arrival, we just spent ten minutes on this.



Herman

Correct, and that's the asterisk on the whole Flic plan. It works beautifully with the phone awake — in hand, screen on. In the pocket with the screen locked, same wall. But it's a much smaller wall than the 8BitDo's, because a Flic is one button you can find blind, on a pocket strap or a collar, without looking.



Corn

One button, seven actions. The arithmetic doesn't obviously work.



Herman

Here's the arithmetic that does: each Flic button gives you three actions — push, double push, and hold. So one Flic is three actions. Two Flics are six. And Android caps paired Bluetooth devices at around seven, so two Flics plus the headset mic plus whatever else leaves him comfortably inside the limit. Two Flics on a strap, six actions, plus the headset's media keys.



Corn

Though "comfortably inside the limit" is a phrase that's never once survived contact with Daniel's desk.



Herman

His desk is its own ecosystem, I grant you. The other thing Flic gets right is the battery — replaceable coin cell, rated twenty-four to thirty-six months. That matters more for a walking setup than people think, because a device you have to charge is a device you'll eventually walk without.



Corn

Twenty-five dollars, wasn't it?



Herman

Thirty-five list. And there's a musician's device worth mentioning because it's the closest thing to what Daniel described wanting — the AirTurn Digit III. Six buttons, handheld, Bluetooth, made for musicians to drive page turns and effects. Launched at thirty-nine dollars in 2017, now forty-five to forty-nine at the retailers that still carry it, discontinued at others.



Corn

So a nine-year-old niche product at a ten-dollar markup on the used-adjacent market.



Herman

Niche products don't get cheaper, Corn. They get rarer and then they get expensive. But it has the same disease as the 8BitDo in one respect — the keystrokes still die at a locked screen. It's a better shape and a better idea with the same operating-system ceiling. Six buttons for a hand that's free is lovely. Daniel's hand is holding nothing, which is lovely too, until a dog appears.



Corn

He records outdoors now. A dog appears every eleven minutes in this city.



Herman

So handheld is conditionally good. Phone in hand or at least awake, Flic or the Digit 3 both work, and the Flic works single-handed and blind. Now. The shoe. I looked for a commercial product — an in-shoe pad, a toe switch, anything that acts as a Bluetooth keyboard from the foot. There isn't one. What exists in-shoe is research instrumentation: sensor insoles with pressure sensors streaming data over Bluetooth. There's a University of Illinois project with an ESP32-S3 and twelve pressure sensors sampling at around a hundred hertz — but that's a data stream, not key presses. Nobody has productised "toe sends the letter R".



Corn

Which means Daniel's idea is a build, not a purchase. ESP32, Bluetooth keyboard library, one force-sensitive resistor under the toes. That's a weekend project, and I've watched him do weekend projects. They're never weekends.



Herman

Here's the physics problem though, and it's the real one. Walking loads the sensor every single step. Every footfall is a pressure event. So the poor device has to distinguish "step" from "command", and the only way it can is if the command is deliberately weird — a toe-curl held for a full second, say, or the sensor placed under the arch-side of the toes where a normal stride doesn't press.



Corn

So to signal "retake" he has to walk with one foot clenched for a second mid-stride.



Herman

Mid-stride, on a downhill, in traffic. The deliberate gesture has to be deliberate enough to be reliable, and the price of reliable is slow. You're spending a second of contortion to save reaching up to a button that takes a quarter of a second.



Corn

A button you can find with your hand. The toe pad you have to find with your toe, in a sock, in the dark, by feel, through leather.



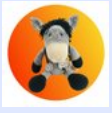
Herman

And there's sweat, and there's the sensor shifting in the shoe over an hour-long walk, and every one of those is a calibration drift problem. The foot pedals that actually work in recording — the AirTurn DUO 500, around a hundred and ten to a hundred and twenty dollars, the PED 500 at about eighty-nine — those are built for a different animal entirely: a stationary musician.



Corn

You keep using that word. Stationary.



Herman

Because it's the whole distinction. Foot input is superb when you're planted — transcriptionists, guitarists, anyone dictating at a desk. The moment you're pacing, the foot is busy doing locomotion and the same toe is now dual-purpose. Hands are free while walking. Feet, very much not.



Corn

The stationary foot pedal is like a great idea with a sign on it that says "do not walk with this".



Herman

Well, that's the question answered, or nearly. But we've got a caller on the line. Dorothy, are you there?

D

Dorothy

Hello dear, it's your mother. I'm here with Dorothy's Kitchen, and today I want to talk about a peeled banana. That's the whole recipe, sweetheart. One banana, peeled. And perhaps a plate, if you're feeling fancy.



Herman

Dorothy, wonderful to have you.



Corn

Hello, Mum.

D

Dorothy

You take the banana in your hand, and you look at it. You look at it for a good long while, because once you start peeling there is no going back. I've seen things go wrong at this stage. I would rather not say how.



Herman

How long a look, Dorothy? Is there a minimum?

D

Dorothy

Then you peel it. Slowly, dear. The skin can turn on you if you rush. And watch the bread while you do it, because the bread always needs watching, even when there is no bread. That is the trick of it.



Corn

Even when there is no bread. Understood, Mum.

D

Dorothy

When it's done, put it on the plate and sit down. Have a cup of tea. And don't forget to eat something proper before your father gets home, sweetheart. Ring me if the butter gives you trouble. Bye bye now.



Corn

She's gone.



Herman

She always is. One banana, peeled. I've written it down, don't worry.



Corn

Where were we. The shoe.



Herman

My verdict on the toe pad — as a build, fun, and I'd cheer him on. As a daily driver for eighty more hours of corpus, it's slower and less reliable than the headset button he already owns. Fun build, worse tool.



Corn

Which takes us to what we'd actually use, which is what he asked for. So let me go first, since I've been pacing around this flat for the last twenty minutes pretending to be him.



Herman

You have not.



Corn

I have. Slowly. It's how I think. Seven actions, and here's my split. First, the app change: he adds next and previous track handlers to the media session — that's a small code change to his own app, and it unlocks the whole headset. Single press is record and stop, double is save and next, triple is retake. Long press he leaves alone; that's the voice assistant.



Herman

And the app already binds everything to keys — space for record, F for save, R for retake, the whole alphabet is spoken for. The media session is the missing translator. It takes a button the system already understands and turns it into an action the app already has.



Corn

Second, the Flic he already owns, clipped to a collar or a pocket strap, for the next tier: push to skip, double push for play last take, hold for pause. Third, stop the run stays on the phone screen. That's the ejector seat.



Herman

With one asterisk: the headset half works with the phone in his pocket, because the media session doesn't care about visibility. The Flic half needs the page awake, so the Flic is for the moments the phone is in his hand.



Corn

That's my answer, and it costs him one Flic he already owns and an afternoon of code.



Herman

Mine's the same shape with one difference, and I want to defend the difference. I'd do the media session work first and live with just the headset for a full week before adding anything. Because I think the honest bet is that three gestures cover most of what he actually does out there: record-stop, save-next, retake. Skip is rarer, and he can always pull the phone out for it.



Corn

You're saying the peripheral question might mostly dissolve.



Herman

I'm saying it might, and the way to find out is nearly free: a few lines of code. The Flic is the backup, not the plan. And if he ever does build the toe pad, I want photos.



Corn

He'll send photos of the second version. The first one will have died in a sock somewhere.



Hilbert

"Peripheral" is the wrong word for it.



Corn

Is it.



Hilbert

It's a controller. A peripheral is a printer. I did a winter recording bird calls for a nature survey — hands full of binoculars and a thermos, no free thumb — and I rigged the sustain pedal off my mother's sewing machine to trigger the recorder. Best controller I ever had. Heavy enough it didn't drift, quiet enough it never spooked a bird, and it never once asked for a firmware update.



Herman

How was it wired in?



Hilbert

Pedal to a gutted USB keyboard, the enter key, paperclip and electrical tape. Three seasons, never missed. Daniel is overcomplicating it. Bolt a pedal to a board and carry the board.



Corn

Carry the board. Down the street.



Hilbert

You pace, Corn, you'd want a board with a strap. Left foot if you're right-handed — non-dominant foot, so the good one stays free for balance. I still use the pedal. Logs my home weather station every morning, same tape on it. Your Bluetooth buttons will be landfill before that tape gives up. Herman bangs on about battery life, thirty-six months — the pedal's been going since before either of you had a desk.



Herman

He's not wrong about the firmware updates.



Corn

He's not wrong about balance either, which is annoying, because we spent ten minutes on the walking problem and the one thing we didn't say is "pick the correct foot".



Hilbert

Left foot. And use the heavy pedal, not the light one — the light one walks with you.



Corn

Well. The board-and-pedal school of mobile computing has a new advocate, and he has tape.



Herman

What I take from Hilbert's winter is the same thing the ESP32 build was circling — a foot switch is only good when the foot has one job. His foot had one job. He was standing in a marsh looking at a finch. Daniel's feet have a job already.



Corn

The finch doesn't care about your calibration window.



Herman

Cutting-room floor, before we land this. One detail I couldn't fit anywhere: how the rest of the speech-recording world handles this. Audiobook narrators, the people who record takes for a living — they use punch-and-roll in Audition and Audacity, driven entirely by keyboard shortcuts. The one controller example I found in that whole literature was a hand-held keypad, a Razer Orbweaver, resting next to the mic. Not one foot pedal. The professionals who could most benefit from feet, because they're seated, still use keys.

Corn

Because keys are where the recovery is. When you fluff a line, the fastest path back is the one your hand already knows.



Herman

Which is the open question I'd leave Daniel with, actually. How much of this friction is really hardware at all? If he took the app's media session all the way — configurable click patterns, double-press means whatever he chooses, all set in a settings page — the peripheral question mostly disappears. He'd be walking around with a headset that is the app. And going forward, as watches and rings and glasses start shipping keyboard modes of their own, the pocketed phone stops being an edge case and becomes the design constraint for every field-recording tool. The button count was never the hard part. The pocket was.



Corn

He's one app update from never buying a gadget again. There's a version of that sentence he'll find upsetting.



Herman

Producing this one, as ever, was Hilbert Flumingtop — pedal intact, tape holding. For more along these lines, there's episode five, Fine-Tuning ASR For Maximal Usability. This has been My Weird Prompts. If the show earns a review from you wherever you listen, it helps other people find us.



Corn

See you soon, everyone.