Una Raspberry Pi 5 en una cabina telefónica

Arquitectura de borde, notas de campo y fiabilidad de un agente de voz físico

Cris Olmedo y Kurt Sommer
Grupo de investigación Divina Machina
Documento de trabajo · Nota técnica
DM-2026-07, versión 1.0 · 30 de octubre de 2026

Resumen. Un agente de voz en un objeto físico necesita decidir qué corre en el dispositivo y qué en la nube. Documentamos la cabina de GAIA 2147 en Barcelona Design Week: una Raspberry Pi 5 que lleva todo lo que se toca, la horquilla, el audio, la luz, la pantalla y el papel, y dos plataformas de voz en la nube, ElevenLabs y Vidiv, con paso automático de una a otra.

Damos la especificación completa, once modos de fallo encontrados en campo con su causa y su corrección, y la fiabilidad registrada en la nube. En Barcelona, el 0,7 % (IC 95 % 0,3–1,9 %) de 543 llamadas terminó con una desconexión anómala y el 0,4 % con un fallo técnico. El primer audio del saludo salió del servidor en 95 ms de mediana con la conexión preparada antes de descolgar. En una ventana de registros de la cabina, el 21 % (IC 95 % 15–29 %) de los colgados agotó una espera de dos segundos en el hilo de audio, un fallo todavía abierto.

La regla que proponemos es simple: el borde se encarga de lo que se toca y de lo que debe pasar en milisegundos; la nube, del lenguaje, la voz y la seguridad, que cambian cada pocos meses.

Palabras clave: computación de borde; Raspberry Pi; agentes de voz; fiabilidad; instalaciones interactivas; análisis de modos de fallo.

1. Introducción

Una cabina telefónica en una feria tiene necesidades sencillas y restricciones duras. Tiene que parecer un teléfono, arrancar sola y funcionar todo el día, sobrevivir al wifi del pabellón y no abrir ningún puerto a internet. La construimos sobre una Raspberry Pi 5: Cortex-A76 de cuatro núcleos a 2,4 GHz, wifi de doble banda, Bluetooth 5.0, dos USB 3 y el conector de 40 pines, con una fuente de 5 V y 5 A [1]. Es suficiente para el mundo físico de la cabina. No es donde debe fabricarse la voz.

Esta nota cubre la versión de Barcelona Design Week. Los registros no documentan el hardware de Sónar+D e IFA; las cifras de fiabilidad de esos eventos son de la nube y no comparan hardware [MED].

2. Arquitectura

La Pi se encarga de todo lo que se puede tocar y del estado de la llamada (figura 1). Pone el tono de marcar, abre la conversación, lleva el audio al auricular y lo trae, controla luz, pantalla e impresora y mantiene reglas y cupos para que nada se dispare. La nube se encarga de la voz, el lenguaje y la seguridad. Siempre trabajamos con dos plataformas, ElevenLabs y Vidiv; si una falla, la cabina pasa a la otra sin intervención [MED]. Un Worker de Cloudflare es el otro extremo de un túnel que abre la cabina, por el que bajan herramientas y ajustes y sube el estado.

3. Modos de fallo

La tabla 2 recoge lo que nos pasó, con el formato de un análisis de modos de fallo. Son observaciones de construcción y operación, no recuentos [MED].

Dos merecen una ecuación. La hoja de datos del WS2812B pide un nivel alto de al menos $0{,}7\,V_{DD}$ [2]:

$$V_{IH,\min} = 0{,}7 \times 5\,\mathrm{V} = 3{,}5\,\mathrm{V} > V_{OH,\mathrm{Pi}} \approx 3{,}3\,\mathrm{V}. \quad (1)$$

Muchas tiras funcionan igual; la nuestra no, y no llegamos a poner el adaptador. La figura 2 muestra la corrección propuesta, un búfer de la familia 74AHCT alimentado a 5 V, cuyo umbral de entrada admite los 3,3 V de la Pi y cuya salida llega a V_{DD} [HIP].

Y la espera del hilo de audio: en una ventana de tres horas y media de registros de la cabina, el 21 % (IC 95 % 15–29 %) de 136 colgados agotó una espera de dos segundos [MED]. Ese recuento sale de los registros de la Pi, que no forman parte del dataset gaia-2147.

4. Fiabilidad

La tabla 3 da la fiabilidad registrada en la nube para los tres eventos [MED]. Barcelona tuvo la menor proporción de desconexiones anómalas, 0,7 % (IC 95 % 0,3–1,9 %), frente al 7,4 % (IC 95 % 5,7–9,6 %) en Sónar+D, cuyo hardware no documentamos. El primer audio del saludo bajó de 287 ms (solo síntesis, Sónar+D) a 95 ms en Barcelona, con la conexión firmada preparada antes de descolgar. Son tiempos de servidor: la Pi y el altavoz añaden su propio retardo, del que solo tenemos lecturas sueltas de entre 0,55 y 0,73 s desde que se abre la conversación hasta el primer sonido.

La carga es real (figura 3): en Barcelona la cabina llegó a 75 llamadas en una hora. Una tarde atendió más de cien llamadas sin caerse; las únicas intervenciones fueron reinicios decididos por nosotros. El coste de nube fue de 0,10 USD por minuto en Barcelona y 0,12 USD en IFA; no calculamos el coste del hardware.

5. Lo que debe hacer el borde

- Todo lo físico y todo lo que debe pasar en milisegundos: horquilla, tonos, luz, pantalla, papel.
- El estado de la llamada, para saber cuándo es seguro imprimir, reiniciar o actualizar.
- El paso entre plataformas: ElevenLabs y Vidiv a la vez, con cambio automático.

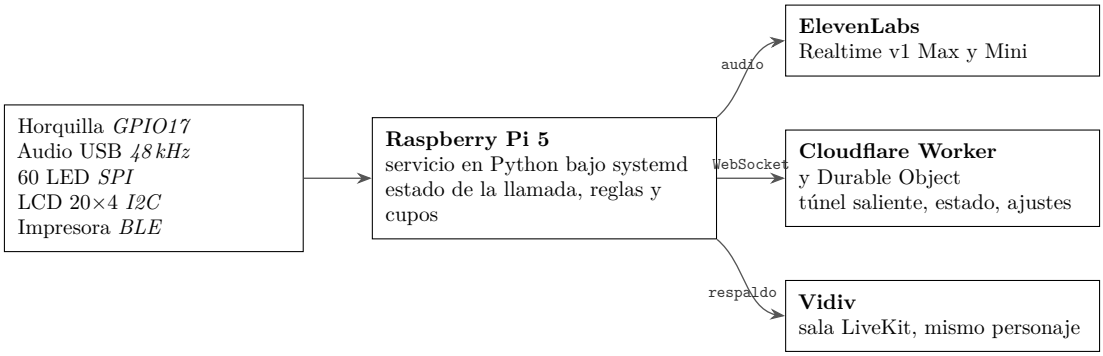


Figura 1. Qué corre dónde. El borde lleva lo físico y el estado; la nube, el lenguaje, la voz y la seguridad. La cabina nunca abre un puerto: todo sale hacia fuera.

Tabla 1. Especificación de la cabina en Barcelona Design Week, extraída de su configuración y código.

Subsistema	Pieza	Interfaz	Ajuste
Ordenador	Raspberry Pi 5		Servicio Python bajo systemd con reinicio automático
Audio	Tarjeta USB	USB, ALSA	48 kHz mono, bloques de 20 ms; mezclador al 98 %; ganancia 3,5; búfer objetivo 40 ms, techo 30 s
Horquilla	Interruptor original	GPIO17	Polaridad invertida por software; antirrebote 120 ms; confirmación 0,4 s
Pantalla	LCD 20×4	I2C 0x27	Murmullos y mensajes cortos con límites por llamada
Luz	60 LED WS2812E	SPI, GPIO10	Brillo 0,35; color de reposo leído de la nube cada 10 s
Papel	Impresora térmica	Bluetooth LE	384 puntos por línea; energía de impresión subida; imprime al colgar
Voz	ElevenLabs Realtime v1	WebSocket	Conexión firmada preparada antes de descolgar
Voz, respaldo	Vidiv	LiveKit	Mismo personaje; paso automático
Enlace	Cloudflare Workers	WebSocket saliente	Estado cada pocos segundos; ajustes cada 10 s
Software	elevenlabs 2.66.0, websockets 16.0		Versiones fijadas; una actualización silenciosa puede deshacer un arreglo

Tabla 2. Modos de fallo encontrados en campo.

Síntoma	Causa	Corrección	Estado
Descolgar no hacía nada	Horquilla con polaridad contraria y rebotes	Lectura invertida, antirrebote y segunda comprobación	Resuelto
Voz entrecortada	Audio a ráfagas y búfer pequeño	Techo del búfer a 30 s, objetivo 40 ms	Resuelto
Volumen reiniciado al arrancar	La tarjeta USB cambiaba de número	Buscarla por nombre	Resuelto
Distorsión con la risa	Ganancia excesiva en el auricular	Mezclador al 98 %, ganancia 3,5	Resuelto
Papel en blanco	Calor de impresión por defecto bajo	Más energía de impresión	Resuelto
Impresora ocupada	Una sola conexión Bluetooth	El servicio la gobierna	Resuelto
Imprimir rompía el audio	Wifi y Bluetooth comparten el módulo de radio	Imprimir solo al colgar	Resuelto
Luz blanca azulada	Datos a 3,3 V frente a un umbral de 3,5 V	Adaptador de nivel y masa común	Abierto
GAIA se contestaba a sí misma	El micrófono recogía el altavoz	Puerta de ruido y micrófono bajo al hablar	Mitigado
Descolgar tardaba unos 2 s más	El hilo de audio a veces no confirmaba pausa o reanudación	Parar sin vaciar y medir cada paso	Abierto
Ajuste de nube tardío	Copia vieja en el objeto del túnel	Consulta de ajustes cada 10 s	Resuelto

Tabla 3. Fiabilidad y saludo por evento, según los registros de la nube. Corchetes, IC 95 % de Wilson o bootstrap.

Evento	Llamadas	Desconexión anómala	Fallo técnico	Saludo (ms)
Sónar+D	701	52 · 7,4 % [5,7; 9,6]	8 · 1,1 % [0,6; 2,2]	287 (TTS)
IFA	274	3 · 1,1 % [0,4; 3,2]	3 · 1,1 % [0,4; 3,2]	179 (e2e)
BDW	543	4 · 0,7 % [0,3; 1,9]	2 · 0,4 % [0,1; 1,3]	95 (e2e)

Desconexión anómala, cierre de WebSocket 1006 o 1011. Fallo técnico, estado fallido sin guardarraíl de seguridad. Saludo, mediana en servidor del primer audio del turno 0; en Sónar+D solo existe el primer byte de síntesis.

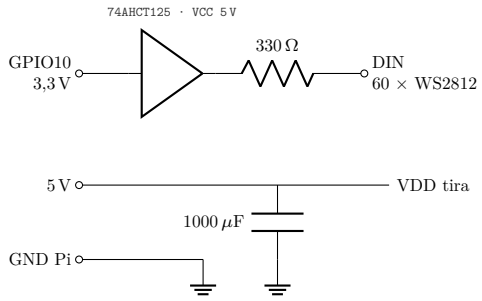


Figura 2. Adaptador de nivel propuesto para la tira de luz, con masa común. Esquema dibujado con circuitikz; no instalado en Barcelona.

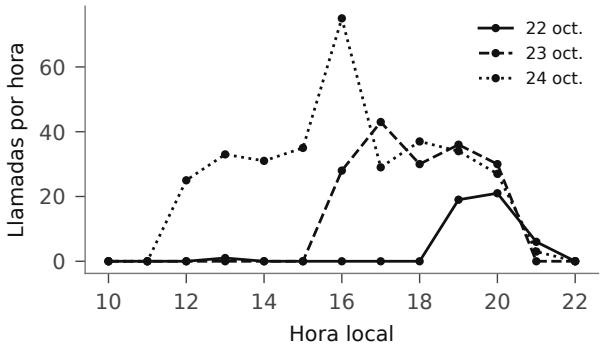


Figura 3. Llamadas por hora en Barcelona Design Week.

- La privacidad por defecto: historial local, subida solo si se activa y con aviso visible.

Y lo que no: reconocimiento, modelo de lenguaje, voz y filtros de seguridad. Cambian cada pocos meses y la cabina no debería necesitar una imagen nueva cuando lo hacen [INT].

6. Amenazas a la validez

Una sola versión en un solo evento. Las notas de campo no están contadas. La fiabilidad de Sónar+D e IFA no corresponde a un hardware documentado. Los tiempos hasta el primer sonido en la cabina son lecturas sueltas. El recuento de esperas agotadas cubre una ventana de registros, no el evento.

7. Recomendaciones para la comunidad

1. Poner un adaptador de nivel en cualquier tira WS2812 alimentada a 5 V desde el primer día.
2. Dar a la impresora su propia radio o pasar a USB; imprimir no debe competir con la voz.
3. Registrar en el dispositivo, cada hora, tiempo en marcha, reinicios, primer sonido y eco.
4. Reiniciar solo con el auricular colgado, comprobándolo dos veces.
5. Fijar versiones de librerías y tener una segunda tarjeta preparada con un único script de instalación.

8. Datos y código

Fiabilidad y tiempos de servidor del dataset gaia-2147 [3]; cifras de `analysis/analyze.py`. Especificación y modos de fallo, de la configuración y los registros de la cabina.

9. Conclusiones

Un ordenador pequeño basta para que un objeto converse si no intenta fabricar la voz. La cabina llevó lo que se toca; la nube, lo que se dice. Lo que falló fue casi siempre físico: polaridades, radios compartidas, voltajes. Esa es la parte que no viene en ninguna ficha de producto.

Firman. Cris Olmedo y Kurt Sommer, Grupo de investigación Divina Machina. *Cómo citar*: Olmedo, C. y Sommer, K. (2026). Una Raspberry Pi 5 en una cabina telefónica. Documento de trabajo DM-2026-07, v 1.0. Grupo de investigación Divina Machina. Divina Machina by Qs Ventures.

Referencias

Fuentes consultadas hasta el 29 de octubre de 2026.

[1] Raspberry Pi Ltd. Raspberry pi 5. Página de producto, 2023. URL https://www.raspberrypi.com/products/raspberry-pi-5/.

[2] Worldsemi. WS2812B. intelligent control LED integrated light source. Hoja de datos, copia de la Universidad de Cambridge. URL https://www.cl.cam.ac.uk/teaching/1516/ECAD+Arch/files/pcb/WS2812B.pdf.

[3] C. Olmedo and K. Sommer. GAIA-2147. descriptor de datos de 1527 conversaciones con un agente de voz en espacios públicos. Documento de trabajo DM-2026-09. Grupo de investigación Divina Machina, 2026.