El primer segundo

Saludo, silencio y continuidad en llamadas a un agente de voz

Cris Olmedo y Kurt Sommer
Grupo de investigación Divina Machina
Documento de trabajo · Estudio observacional
DM-2026-04, versión 1.0 · 30 de octubre de 2026

Resumen. Un agente de voz en un teléfono público tiene un segundo para demostrar que la línea está viva. Estudiamos tres aperturas del mismo personaje, GAIA 2147, en 543 llamadas de Barcelona Design Week: ningún saludo, un saludo largo y fijo, y un saludo corto elegido al azar entre 8 variantes.

Sin saludo, el 71 % de las llamadas duró menos de diez segundos y en el 62 % el visitante no dijo nada. Con el saludo corto, esas cifras fueron el 7 % y el 6 %, frente al 14 % y el 24 % del saludo largo. Un análisis de supervivencia que trata como censuradas las llamadas cortadas por el límite de tiempo confirma el orden de las tres curvas. En el mismo evento, cuando la primera respuesta del agente llegó en dos segundos o menos, el visitante volvió a hablar en el 96 % de las llamadas, frente al 81 % cuando tardó más.

Leemos el resultado con la teoría de las aperturas telefónicas. Quien descuelga un teléfono público es quien llama, y espera que el otro lado hable primero. Las tres aperturas coinciden con cambios de prompt y de hora; el diseño es observacional y lo tratamos como tal.

Palabras clave: apertura de llamada; saludo; agentes de voz; análisis de supervivencia; toma de turnos; diseño conversacional.

1. Introducción

Las aperturas de las llamadas son de las secuencias mejor estudiadas de la conversación. Schegloff mostró que la primera intervención de una llamada responde a la llamada misma, el timbre, y que quien contesta habla primero [1]. Los agentes de voz heredan esa gramática sin haberla elegido. GAIA 2147 es un personaje que contesta en una cabina telefónica en espacios públicos. El visitante descuelga para ver qué pasa, sin tarea ni instrucciones.

En Barcelona Design Week cambiamos la apertura dos veces con la cabina abierta. Eso deja tres grupos con el mismo hardware, la misma pila de voz y el mismo evento. No es un experimento aleatorizado. Es la comparación más limpia que permite iterar en público [2].

1.1. Preguntas

1. ¿Cómo cambian el abandono temprano y el silencio del visitante con la apertura?
2. ¿Cambia la duración de la llamada cuando se trata el límite de tiempo como censura?
3. ¿Se asocia la rapidez de la primera respuesta con que el visitante vuelva a hablar?

2. Datos y método

2.1. Grupos

La clase de saludo se asigna comparando la primera frase de GAIA, si suena en los dos primeros segundos, con nuestros propios saludos [DEF]. Los grupos son *sin saludo* (24 llamadas), en el que GAIA espera a que hable el visitante; *saludo largo fijo* en español (313), que preguntaba si había alguien y cómo se llamaba; y *saludo corto variable* (140), una de 8 variantes como *¿Sí…?*, *Hello…?* o *Hallo…?*, sin repetir la anterior.

2.2. Variables

Llamada corta, menos de 10 s. *Llamada muda*, sin ningún turno transcrito del visitante. *Continuidad*, que el visitante produzca al menos dos turnos, entre las llamadas en que ya produjo uno. *Primera respuesta*, el primer turno de GAIA con latencia registrada tras el saludo [DEF].

2.3. Estadística

Proporciones con intervalos de Wilson [3] y diferencias con intervalos de Newcombe [4], contrastadas con la prueba exacta de Fisher y corregidas por Holm [5]. Las llamadas que alcanzan el límite de tiempo no terminan por decisión del visitante; tratarlas como duraciones completas sesga la mediana hacia abajo. Por eso estimamos la función de supervivencia de la llamada con Kaplan-Meier [6], con censura en el límite,

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right), \tag{1}$$

donde d_i es el número de llamadas que terminan en t_i y n_i las que siguen abiertas. Las curvas se comparan con la prueba de rangos logarítmicos.

3. Resultados

3.1. Abandono y silencio

La tabla 1 y la figura 1 resumen los tres grupos [MED]. El silencio inicial fue la peor apertura con diferencia. La mitad de esas llamadas duró cero segundos: la gente descolgaba, no oía nada y colgaba.

Frente al saludo largo, no saludar aumentó las llamadas cortas en 56 pp (IC 95 % 36–71 pp) y las mudas en 39 pp (IC 95 % 18–55 pp). El saludo corto las redujo en −7 pp (IC 95 % −13 a −1 pp) y −18 pp (IC 95 % −23 a −11 pp) respectivamente (tabla 2). La reducción de llamadas mudas es la más grande y la más precisa: un saludo de una palabra le devuelve el turno al visitante.

Tabla 1. Resultados por apertura en Barcelona Design Week.

Saludo	n	Corta < 10 s	Muda	Mediana (s)	KM (s)	Límite
Sin saludo	24	71 % [51; 85]	62 % [43; 79]	0 [0; 9]	0	0 %
Saludo largo fijo	313	14 % [11; 19]	24 % [20; 29]	79 [63; 88]	79	32 %
Saludo corto variable	140	7 % [4; 13]	6 % [3; 12]	116 [94; 125]	116	48 %

Corchetes, IC 95 % (Wilson para proporciones, bootstrap para medianas). KM, mediana de Kaplan-Meier con las llamadas cortadas por el límite tratadas como censuradas.

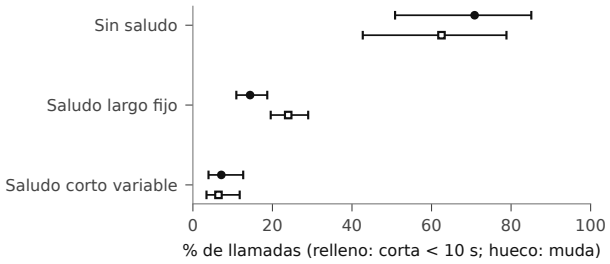


Figura 1. Proporción de llamadas cortas y mudas por apertura, con IC 95 % de Wilson.

Tabla 2. Diferencias de riesgo frente al saludo largo, en puntos porcentuales, con IC de Newcombe y p de Fisher ajustada por Holm.

Comparación	Resultado	Diferencia [IC 95 %]	p Holm
Sin saludo frente a largo	Llamada corta	56 pp [36; 71]	<0,001
Sin saludo frente a largo	Llamada muda	39 pp [18; 55]	<0,001
Saludo corto variable frente a largo	Llamada corta	−7 pp [−13; −1]	0,03
Saludo corto variable frente a largo	Llamada muda	−18 pp [−23; −11]	<0,001

3.2. Duración con censura

Con Kaplan-Meier, la mediana de duración fue 79 s con el saludo largo y 116 s con el corto; sin saludo fue 0 s (figura 2). Las curvas difieren frente al saludo largo tanto sin saludo (log-rank $\chi^2 = 112{,}6$, $p < 0{,}001$) como con el saludo corto ($\chi^2 = 11{,}7$, $p < 0{,}001$). El 48 % de las llamadas con saludo corto llegó al límite de tiempo, frente al 32 % con el largo: la duración real que la gente quería quedarse está por encima de lo que se observa.

3.3. Mismas horas

El saludo corto se usó en la tarde y noche del 24 de octubre; el largo, en los días anteriores y en las primeras horas de esa tarde. Restringido al 24 desde las 14 h, las llamadas cortas fueron el 7 % (IC 95 % 4–13 %) con el saludo corto (140 llamadas) y el 24 % (IC 95 % 17–33 %) con el largo (109), una diferencia de −17 pp (IC 95 % −26 a −8 pp). La dirección se mantiene en horas adyacentes del mismo día [MED], aunque no en las mismas horas.

3.4. Variantes

Las 8 variantes del saludo corto tienen entre 14 y 26 llamadas cada una (tabla 3). Hay heterogeneidad (chi cuadrado de homogeneidad, $p = 0{,}03$), concentrada en *¿Diga…?*, pero los intervalos se solapan casi por completo. No ordenamos las variantes; haría falta un estudio con más llamadas por variante [INT].

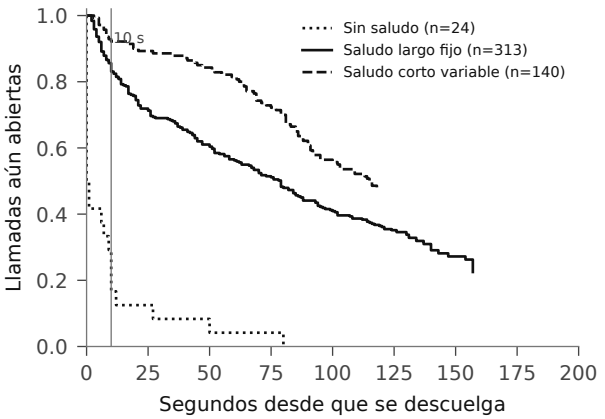


Figura 2. Supervivencia de la llamada por apertura (Kaplan-Meier). Las llamadas cortadas por el límite de tiempo se tratan como censuradas.

Tabla 3. Variantes del saludo corto. Corchetes, IC 95 % de Wilson.

Variante	n	Corta [IC 95 %]	Mediana (s)
¿Sí…?	26	4 % [1; 19]	92
¿Hola…? ¿Me oyes?	19	16 % [6; 38]	88
¿Hola…?	18	6 % [1; 26]	116
Hallo…?	17	0 % [0; 18]	125
¿Alguien…?	16	6 % [1; 28]	108
Hello…?	16	0 % [0; 19]	125
¿Diga…?	14	29 % [12; 55]	68
¿Hola? ¿Hay alguien ahí?	14	0 % [0; 22]	125

3.5. La primera respuesta

En Barcelona Design Week, cuando la primera respuesta de GAIA llegó en dos segundos o menos, el visitante volvió a hablar en el 96 % (IC 95 % 94–98 %) de las llamadas (359); cuando tardó más, en el 81 % (IC 95 % 69–89 %) (53), una diferencia de 15 pp (IC 95 % 6–28 pp) (Fisher, $p < 0{,}001$). En IFA casi todas las primeras respuestas fueron lentas; con 22 llamadas rápidas la diferencia es imprecisa, 8 pp (IC 95 % −14 a 22 pp). La latencia por pila se analiza en DM-2026-03.

4. Discusión

Quien descuelga es quien llama. En la secuencia que describe Schegloff, el timbre convoca y quien contesta habla primero [1]. En una cabina pública el visitante

descuelga para hacer una llamada, no para atenderla, y espera tono o voz del otro lado. Cuando GAIA calla, el visitante oye una línea muerta [INT]. El saludo corto es el turno del que contesta, el *¿Diga?* de siempre, y deja al visitante el siguiente turno. El saludo largo hace lo mismo pero gasta los primeros segundos en la máquina.

El silencio al teléfono no es copresencia. Hoey distingue los silencios permitidos entre personas presentes de los lapsos que se notan como ausencia [7]. Al teléfono no hay un cuerpo que sostenga el silencio; por encima de un segundo, el silencio ya significa [8]. La asociación entre primera respuesta lenta y abandono apunta en la misma dirección, con la salvedad de que una pregunta difícil puede alargar la respuesta y acortar la llamada a la vez [INT].

5. Amenazas a la validez

- *Confusión temporal.* Los grupos ocupan franjas distintas y coinciden con cambios de prompt descritos en DM-2026-05. El análisis de mismas horas reduce, no elimina, esa confusión.
- *Sin saludo* tiene 24 llamadas; su efecto es grande y su intervalo, ancho.
- *Clasificación.* La clase de saludo sale de comparar texto del agente con nuestros saludos; las aperturas de otra clase quedan fuera.
- *Continuidad* se mide solo entre quienes ya hablaron una vez, una población seleccionada.

6. Recomendaciones para la comunidad

1. No empezar nunca con silencio en un canal de voz sin imagen.
2. Abrir como quien contesta: una palabra corta, variable, que devuelva el turno.
3. Medir la apertura con tres cifras: llamadas cortas, llamadas mudas y supervivencia con censura en el límite.
4. Medir aparte la latencia de la primera respuesta; es la que decide si hay segunda.
5. Para comparar variantes, planificar al menos varios cientos de llamadas por variante; con decenas solo se ven efectos enormes.

7. Datos y código

Dataset gaia-2147 [2]; cifras generadas por `analysis/analyze.py`. Siempre trabajamos con ElevenLabs y Vidiv; estos registros son los del agente de ElevenLabs. No se guardó texto de los visitantes.

8. Conclusiones

El primer segundo de una llamada es una decisión de diseño. Callar convirtió siete de cada diez llamadas en un descolgar y colgar. Una palabra corta, dicha como quien contesta, llevó las llamadas mudas al 6 %. El personaje empieza antes de decir quién es.

Firman. Cris Olmedo y Kurt Sommer, Grupo de investigación Divina Machina. *Cómo citar*: Olmedo, C. y Sommer, K. (2026). El primer segundo. Documento de trabajo DM-2026-04, v 1.0. Grupo de investigación Divina Machina. Divina Machina by Qs Ventures.

Referencias

Fuentes consultadas hasta el 29 de octubre de 2026.

[1] E. A. Schegloff. Sequencing in conversational openings. *American Anthropologist*, 70(6):1075–1095, 1968. doi: 10.1525/aa.1968.70.6.02a00030.

[2] 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.

[3] E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927. doi: 10.1080/01621459.1927.10502953.

[4] R. G. Newcombe. Interval estimation for the difference between independent proportions. comparison of eleven methods. *Statistics in Medicine*, 17(8):873–890, 1998.

[5] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[6] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958. doi: 10.1080/01621459.1958.10501452.

[7] E. M. Hoey. *When Conversation Lapses. The Public Accountability of Silent Copresence*. Oxford University Press, 2020. URL https://academic.oup.com/book/40563.

[8] G. Jefferson. Preliminary notes on a possible metric which provides for a standard maximum silence of approximately one second in conversation. In D. Roger and P. Bull, editors, *Conversation. An Interdisciplinary Perspective*, pages 166–196. Multilingual Matters, 1989. URL https://emcawiki.net/Jefferson1989a.