The first second

Greeting, silence and continuity in calls to a voice agent

Cris Olmedo and Kurt Sommer
Divina Machina Research Group
Working paper · Observational study
DM-2026-04, version 1.0 · 30 October 2026

Abstract. A voice agent on a public telephone has one second to show that the line is alive. We study three openings by the same character, GAIA 2147, across 543 calls at Barcelona Design Week: no greeting, a long fixed greeting, and a short greeting drawn at random from 8 variants.

With no greeting, 71 % of calls lasted under ten seconds and in 62 % the visitor said nothing. With the short greeting those figures were 7 % and 6 %, against 14 % and 24 % for the long greeting. A survival analysis that treats calls cut by the time limit as censored confirms the order of the three curves. At the same event, when the agent's first response arrived within two seconds the visitor spoke again in 96 % of calls, against 81 % when it took longer.

We read the result through the theory of telephone openings. Whoever picks up a public phone is the caller, and expects the other end to speak first. The three openings coincide with changes of prompt and time of day; the design is observational and we treat it as such.

Keywords: call opening; greeting; voice agents; survival analysis; turn-taking; conversation design.

1. Introduction

Call openings are among the best-studied sequences in conversation. Schegloff showed that the first utterance of a call answers the summons itself, the ring, and that the answerer speaks first [1]. Voice agents inherit that grammar without having chosen it. GAIA 2147 is a character who answers in a telephone booth in public spaces. The visitor picks up to see what happens, with no task and no instructions.

At Barcelona Design Week we changed the opening twice with the booth open. That leaves three groups with the same hardware, the same voice stack and the same event. It is not a randomised experiment. It is the cleanest comparison that iterating in public allows [2].

1.1. Questions

1. How do early abandonment and visitor silence change with the opening?
2. Does call duration change when the time limit is treated as censoring?
3. Is the speed of the first response associated with the visitor speaking again?

2. Data and method

2.1. Groups

The greeting class is assigned by matching GAIA's first sentence, if it plays within the first two seconds, against our own greetings [DEF]. The groups are *no greeting* (24 calls), in which GAIA waits for the visitor to speak; *long fixed greeting* in Spanish (313), which asked whether anyone was there and what their name was; and *short varied greeting* (140), one of 8 variants such as *¿Sí…?*, *Hello…?* or *Hallo…?*, never repeating the previous one.

2.2. Variables

Short call, under 10 s. *Mute call*, with no transcribed visitor turn. *Continuity*, the visitor producing at least two turns, among calls in which they already produced one. *First response*, GAIA's first turn with recorded latency after the greeting [DEF].

2.3. Statistics

Proportions with Wilson intervals [3] and differences with Newcombe intervals [4], tested with Fisher's exact test and corrected with Holm [5]. Calls that reach the time limit do not end by the visitor's choice; treating them as complete durations biases the median downwards. We therefore estimate the call survival function with Kaplan-Meier [6], censoring at the limit,

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

where d_i is the number of calls ending at t_i and n_i those still open. Curves are compared with the log-rank test.

3. Results

3.1. Abandonment and silence

Table 1 and figure 1 summarise the three groups [MED]. Initial silence was by far the worst opening. Half of those calls lasted zero seconds: people picked up, heard nothing and hung up.

Compared with the long greeting, no greeting increased short calls by 56 pp (95 % CI 36–71 pp) and mute calls by 39 pp (95 % CI 18–55 pp). The short greeting reduced them by −7 pp (95 % CI −13 to −1 pp) and −18 pp (95 % CI −23 to −11 pp) respectively (table 2). The reduction in mute calls is the largest and the most precise: a one-word greeting hands the turn back to the visitor.

3.2. Duration with censoring

With Kaplan-Meier, median duration was 79 s with the long greeting and 116 s with the short one; with

Table 1. Outcomes by opening at Barcelona Design Week.

Greeting	n	Short < 10 s	Mute	Median (s)	KM (s)	Limit
No greeting	24	71 % [51; 85]	62 % [43; 79]	0 [0; 9]	0	0 %
Long fixed greeting	313	14 % [11; 19]	24 % [20; 29]	79 [63; 88]	79	32 %
Short varied greeting	140	7 % [4; 13]	6 % [3; 12]	116 [94; 125]	116	48 %

Brackets, 95 % CI (Wilson for proportions, bootstrap for medians). KM, Kaplan-Meier median with calls cut by the time limit treated as censored.

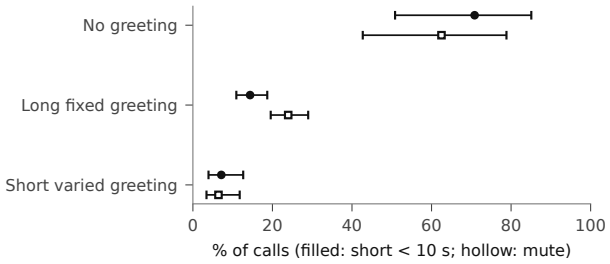


Figure 1. Share of short and mute calls by opening, with Wilson 95 % CI.

Table 2. Risk differences against the long greeting, in percentage points, with Newcombe CI and Fisher p adjusted by Holm.

Comparison	Outcome	Difference [95 % CI]	Holm p
No greeting vs long	Short call	56 pp [36; 71]	<0.001
No greeting vs long	Mute call	39 pp [18; 55]	<0.001
Short varied greeting vs long	Short call	−7 pp [−13; −1]	0.03
Short varied greeting vs long	Mute call	−18 pp [−23; −11]	<0.001

no greeting it was 0 s (figure 2). The curves differ from the long greeting both for no greeting (log-rank $\chi^2 = 112.6$, $p < 0.001$) and for the short greeting ($\chi^2 = 11.7$, $p < 0.001$). 48 % of calls with the short greeting reached the time limit, against 32 % with the long one: how long people actually wanted to stay is above what is observed.

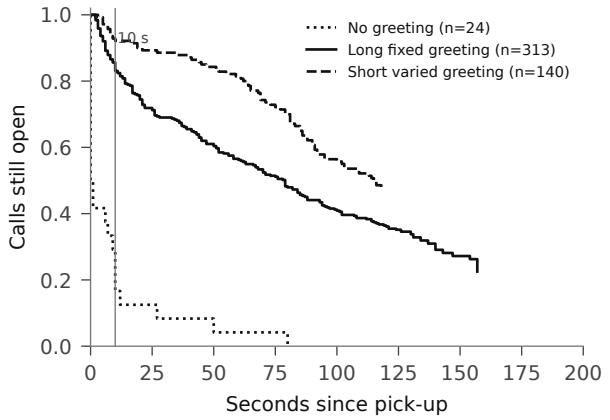


Figure 2. Call survival by opening (Kaplan-Meier). Calls cut by the time limit are treated as censored.

Table 3. Variants of the short greeting. Brackets, Wilson 95 % CI.

Variant	n	Short [95 % CI]	Median (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

3.3. Same hours

The short greeting was used on the afternoon and evening of 24 October; the long one on the previous days and in the first hours of that afternoon. Restricted to the 24th from 14:00, short calls were 7 % (95 % CI 4–13 %) with the short greeting (140 calls) and 24 % (95 % CI 17–33 %) with the long one (109), a difference of −17 pp (95 % CI −26 to −8 pp). The direction holds in adjacent hours of the same day [MED], though not in the same hours.

3.4. Variants

The 8 variants of the short greeting have between 14 and 26 calls each (table 3). There is heterogeneity (chi-square test of homogeneity, $p = 0.03$), concentrated in *¿Diga…?*, but the intervals overlap almost entirely. We do not rank the variants; that would need a study with more calls per variant [INT].

3.5. The first response

At Barcelona Design Week, when GAIA's first response arrived within two seconds, the visitor spoke again in 96 % (95 % CI 94–98 %) of calls (359); when it took longer, in 81 % (95 % CI 69–89 %) (53), a difference of 15 pp (95 % CI 6–28 pp) (Fisher, $p < 0.001$). At IFA almost every first response was slow; with 22 fast calls the difference is imprecise, 8 pp (95 % CI −14 to 22 pp). Latency by stack is analysed in DM-2026-03.

4. Discussion

Whoever picks up is the caller. In the sequence Schegloff describes, the ring summons and the answerer speaks first [1]. In a public booth the visitor picks up to make a call, not to take one, and expects a tone or a voice from the other end. When GAIA stays silent, the visitor hears a dead line [INT]. The short greeting is the answerer's turn, the familiar *Hello?*, and leaves

the next turn to the visitor. The long greeting does the same but spends the first seconds on the machine.

Silence on the phone is not copresence. Hoey distinguishes permitted silences between copresent people from lapses noticed as absence [7]. On the phone there is no body to hold the silence; beyond a second, silence already carries meaning [8]. The association between a slow first response and abandonment points the same way, with the caveat that a difficult question can both lengthen the response and shorten the call [INT].

5. Threats to validity

- *Temporal confounding.* The groups occupy different time slots and coincide with prompt changes described in DM-2026-05. The same-hours analysis reduces this confounding but does not remove it.
- *No greeting* has 24 calls; its effect is large and its interval wide.
- *Classification.* The greeting class comes from matching agent text to our greetings; openings of other classes are excluded.
- *Continuity* is measured only among those who already spoke once, a selected population.

6. Recommendations for the community

1. Never open with silence on a voice channel without image.
2. Open like the answerer: one short, varied word that hands the turn back.
3. Measure the opening with three figures: short calls, mute calls and survival with censoring at the limit.
4. Measure first-response latency separately; it decides whether there is a second one.
5. To compare variants, plan for at least several hundred calls per variant; with dozens only huge effects show.

7. Data and code

gaia-2147 dataset [2]; figures generated by `analysis/analyze.py`. We always work with ElevenLabs and Vidiv; these logs are those of the ElevenLabs agent. No visitor text was stored.

8. Conclusions

The first second of a call is a design decision. Silence turned seven in ten calls into a pick-up and hang-up. One short word, said like someone answering, brought mute calls down to 6 %. The character begins before it says who it is.

Signed. Cris Olmedo and Kurt Sommer, Divina Machina Research Group. *How to cite*: Olmedo, C. and Sommer, K. (2026). The first second. Working paper DM-2026-04, v 1.0. Divina Machina Research Group. Divina Machina by Qs Ventures.

References

Sources consulted up to 29 October 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. data descriptor of 1527 conversations with a voice agent in public spaces. Working paper DM-2026-09. Divina Machina Research Group, 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.