



Voice Agent QA Sprint

Executive QA Report — Pilot Campaign

Voice agent under test: « Atlas » (demo) · **Language:** French with North African accents · **Campaign:** Week 22 · **Volume:** 320 calls / 74 scenarios

320 Calls tested	74 Scenarios executed	86% Task completion	27 Confirmed defects
---------------------	--------------------------	------------------------	-------------------------

Sample report — demonstration data only. All figures, defect identifiers and findings in this document are illustrative. They show the structure, depth and presentation of the report you receive after a real Voice Agent QA Sprint.

Rapport exemple — données de démonstration uniquement. Tous les chiffres et constats de ce document sont illustratifs.

Contents

- 1. Campaign overview
- 2. Key performance indicators
- 3. Results by language and accent
- 4. Defect analysis
- 5. Defect log excerpt
- 6. Priority recommendations

1. Campaign overview

The Sprint tested the voice agent « Atlas » with real human callers across French (France), North African French, Tunisian Arabic and French–Arabic code-switching scenarios. Callers followed scripted and natural scenarios: appointment booking, billing enquiry, complaint handling, technical assistance and sales qualification, plus edge cases — barge-in, silence, background noise, objections and transfer-to-human.

Attribute	Value	Attribute	Value
Language	French (+ AR dialects)	Scenarios	74
Voice agent	« Atlas » (demo)	Human calls	320
Calls per scenario	2–6	Callers	24 (qualified native speakers)
Duration	1–8 min (mean 3.6 min)	Recording	Dual-channel, 16 kHz WAV
Defects logged	27 (3 critical / 9 major / 15 minor)	Review meeting	Day 5 after delivery

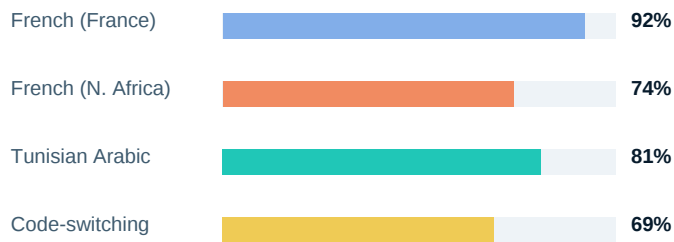
Method: every call was scored on task completion, intent recognition, response quality, interruption handling and linguistic/cultural adequacy. Each defect was reproduced by at least two callers before being logged.

2. Key performance indicators

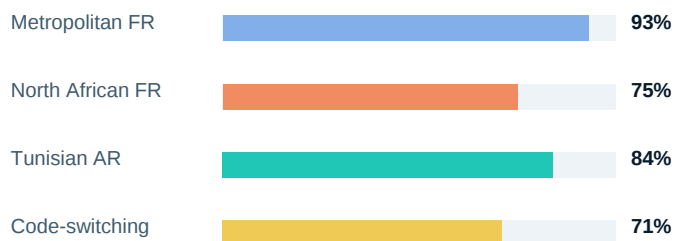
KPI	Result	Pilot target	Verdict
Task completion (overall)	86%	80%	▲ Achieved
Task completion — Metropolitan French	92%	85%	▲ Achieved
Task completion — North African French	74%	80%	▼ Missed (-18 pp vs. Metro FR)
Task completion — Tunisian Arabic	81%	80%	▲ Achieved
Task completion — code-switching calls	69%	75%	▼ Missed
Fallback rate	12%	< 15%	▲ Within range
Human transfer success	87%	80%	▲ Achieved
Direct resolution without fallback	88%	85%	▲ Achieved
Average duration (solved calls)	3.6 min	2–5 min	▲ Within range

3. Results by language and accent

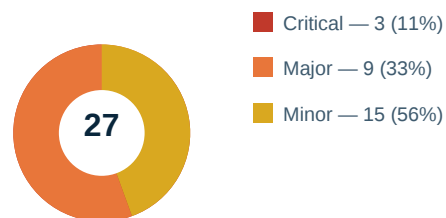
Task completion by language



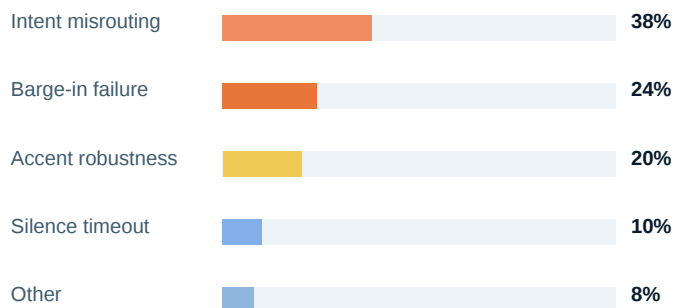
Task completion by speaker accent



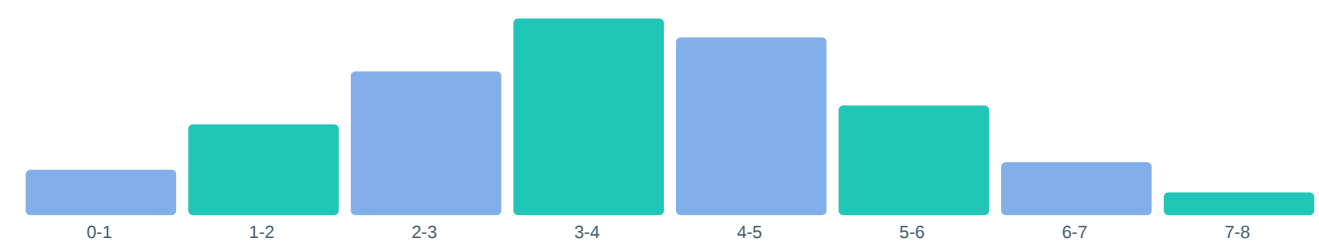
4. Defect analysis



Failure categories (% of failed scenarios)



Conversation duration distribution (minutes)



5. Fallback and transfer behavior

Behavior	Rate	Comment
Direct resolution (no fallback)	88%	Within pilot target (≥85%)
Fallback triggered	12%	Within pilot target (<15%)
Transfer-to-human success	87%	Transfers accepted by human queue within 60 s
Abandoned after fallback	4%	2% above target — investigate greeting after transfer

6. Defect log excerpt (reproducible, de-identified)

Each defect below was observed by at least two independent callers and includes the reproducing scenario and utterance category. Full log with audio references is delivered separately under NDA.

ID	Sev.	Category	Scenario	Observed behavior	Repro rate
D-01	Critical	Intent misrouting	Complaint / code-switched opening	« J'ai un problème avec ma facture » → routed to sales intent. Misrouting reproduces whenever the opening utterance mixes French and Darija.	8/8
D-06	Critical	Intent misrouting	Billing / Arabic numerals	Amount read in Tunisian Arabic digits (« khamsa o tlatin ») → wrong slot value (53 vs 35).	6/6
D-09	Critical	Accent robustness	Tech support / North African French	Word-final consonant drop (« connecté » → « connecté? ») triggers clarification loop; task abandoned in 5/8 calls.	5/8
D-12	Major	Barge-in failure	Sales qualification	Agent keeps speaking 2.1–4.4 s after caller interruption; caller repeated request in 6/9 occurrences.	6/9
D-15	Major	Silence timeout	Appointment booking	Silence > 1.8 s after a request for a date fires the exit prompt, even with confirmations pending.	4/7
D-18	Major	Objection handling	Complaint handling	« C'est inadmissible » → apology only, no resolution path; escalation proposed only after 2 repetitions.	5/7
D-21	Minor	Linguistic quality	All / opening	Asymmetric « vous » / « tu » switching during the same call; inconsistent register in 15% of calls.	15%
D-24	Minor	Response quality	Billing	Amounts repeated too fast; caller asked to repeat in 6/9 occurrences.	6/9

7. Priority recommendations

P1 — Critical — Fix intent misrouting on code-switched openings (D-01, D-06). Add French–Darija mixed-language test suite to regression before next release.

P2 — Critical — Retrain accent model on North African French; add word-final consonant and digit pronunciation variants to the lexicon (D-09, D-21).

P3 — Major — Implement barge-in rejection with 300 ms grace and split-channel echo cancellation; re-test with the S-07 noise profile (D-12).

P4 — Major — Extend silence timeout from 1.8 s to 3.2 s when a confirmation is pending (D-15).

P5 — Minor — Standardize register policy (vous/tu) per scenario; add post-processing normalization layer (D-21).

Next steps: optional regression Sprint on the same scenario set (re-run recommended after fixing P1–P3), or a 20-Call Discovery Benchmark on the second voice agent before its pilot.

Ceci est un rapport exemple. Pour un pilote réel : contact.synergycall@gmail.com · +33 6 20 23 15 96 · +216 57 01 09 04