

SpeechLess: Micro-utterance with Personalized Spatial Memory-aware Assistant in Everyday Augmented Reality

Yoonsang Kim
Stony Brook University

Yalong Yang
Georgia Institute of Technology

Devshree Jadeja
Stony Brook University

Divyansh Pradhan
Stony Brook University

Arie E. Kaufman
Stony Brook University

SUPPLEMENTARY MATERIALS

On this page, we illustrate any details that can further support the claims of our main manuscript. The following details are supporting arguments of our research. Optional, but can be found valuable.

1 Implementation Settings

SpeechLess is developed with Unity AR Foundation and was deployed on both Meta Quest3 HMD (Passthrough AR mode), and hand-held AR platforms (Android, iOS). The client app of SpeechLess, is connected to our custom Python server hosted on Google Cloud Platform, to offload compute-heavy processes including voice/image analysis and LLM interactions. Also, for Quest3, we pair with an auxiliary smart device to provide accurate GPS location of the user. The GPS location is converted to textual representation via OpenStreetMap Nominatim [4] for a better LLM comprehension. The storing and recalling of memories are crucial internal operations of SpeechLess. We leverage two multimodal LLM models: GPT-4.1 [5] and Gemini 2.5-Flash [1], for generating memories, and textual representations of scene/activity, and referents. Semantic embeddings model (text-embedding-3-large) was used for initial memory retrieval with a graph-based approximate neighbor search algorithm, HNSW [3]. The resulting memory candidates were re-ranked using a late-interaction model (ColBERTv2). For user interactions, we adopt head-gaze raycast metaphor – users center their head-view towards the object-of-interest – and optionally use hand-pointing gesture to direct the object-of-interest [2]. OpenAI Whisper and TTS API (gpt-4o-mini-tts) were used to transcribe and generate audio response.

2 Study Design

The following content is used to provide a guideline to the participants of our study (**Study A**), while allowing us to limit the scope of the study within a controlled/predictable settings. We use the concept ‘Persona’ and assign “Personal Spatial Memory” to each, across three varying test environment (Cafe/Bistro, Office Break-Room, Office Desk-**Fig. 1**). Each Seeded question involves, the referent it is anchored to, and the question the person (of the assigned Persona) queried earlier that is saved in the person’s Spatial Memory. The personas were randomly assigned to each participant.

2.1 Persona 1: Creative Entrepreneur

Persona/Role: You always think about an idea, and work. Every action inspires a new start-up idea, research, and you tend to be a workaholic. Your life revolves around coming up with a next generation product and content, and you think a lot during your everyday routine, and leave personal thought-notes.

Seeded Questions per Location:

Format: (object-of-inquiry) "question"

Office Desk

[1] (calendar) "What deadline did I set on the calendar?"
[2] (left monitor) "What project name is on the left screen?"

[3] (right monitor) "What dashboard metric did I track?"
[4] (keyboard) "Which shortcut key did I save for edits?"
[5] (mouse) "What task did I flag as blocked?"
[6] (desk lamp) "What focus reminder did I attach to the lamp?"
[7] (Reese’s mug) "What idea did I note while using this mug?"
[8] (glasses case) "What product name variant did I store?"
[9] (badge lanyards) "Which conference lead did I note?"
[10] (drawer) "What hiring priority did I park here?"
[11] (water bottle) "What ‘hydrate before pitch’ rule did I save?"
[12] (post-it) "What two-word demo mantra did I write?"
[13] (keyboard) "Which macro did I record for versioning?"
[14] (calendar) "What milestone did I mark as investor-ready?"
[15] (desk surface) "What competitor teardown did I list?"

Office Breakroom or Pantry

[1] (espresso machine) "Which meeting did I pair with espresso?"
[2] (French press) "What deep-work block did I plan here?"
[3] (coffee menu) "Which price inspired my tier plan?"
[4] (open hours sign) "What standup time did I pick?"
[5] (paper cups) "What tagline did I write on a cup?"
[6] (plant) "What creativity cue did I save?"
[7] (fridge) "What ‘brain fuel’ snack did I choose?"
[8] (water pitcher) "What refill cadence did I note?"
[9] (logo sign) "What brand color pairing did I like?"
[10] (bathroom sign) "What wayfinding copy did I prefer?"
[11] (countertop) "What guerilla-interview plan did I jot?"
[12] (kettle) "What ‘warm-start’ ritual did I set?"
[13] (espresso cups) "What 30-sec pitch did I rehearse?"
[14] (outlet area) "What charging-perk idea did I log?"
[15] (grinder) "What coffee-named codename did I pick?"

Cafe or Bistro

[1] (menu poster) "Which lunch special sparked a bundle idea?"
[2] (garlic sauce) "What label text did I reuse?"
[3] (wasabi soy) "What ‘spicy add-on’ name did I note?"
[4] (water bottle) "What reset phrase did I write?"
[5] (cork jar) "Which milestone did I tie to ‘corks’?"
[6] (bowls) "What minimal-UI note did I save?"
[7] (fork \& spoon) "What dual-mode feature did I map?"
[8] (glasses) "What transparency rule did I set?"
[9] (napkins) "What cleanup step did I add?"
[10] (tablecloth grid) "What layout grid did I choose?"
[11] (Asian Food logo) "What color duo did I prefer?"
[12] (bathroom sign) "What onboarding arrow text did I like?"
[13] (plant) "What growth-loop note did I write?"
[14] (right seat) "Which persona sits on the right?"
[15] (left seat) "Which stakeholder sits on the left?"

2.2 Persona 2: Social Extrovert

Persona/Role: An outgoing extrovert who loves socializing, meeting friends, and new people. Always recalling moments with friends, and seeking for new experiences. The relationship between people and having a good time with them is your highest value.

Seeded Questions per Location:

Format: [index] (object-of-inquiry) "question"



Figure 1: The three user interaction spaces of *Study A*: **(A)** Cafe/Bistro **(B)** Office BreakRoom **(C)** Office Desk.

Office Desk

[1] (calendar) "Whose birthday did I save?"
 [2] (left monitor) "Whose message did I open here?"
 [3] (right monitor) "What playlist did I queue for co-work?"
 [4] (Reese's mug) "Who gifted me this mug?"
 [5] (clamp) "Who am I meeting after work?"
 [6] (badge lanyards) "Who did I promise to DM?"
 [7] (water bottle) "Who joins me on coffee walks?"
 [8] (glasses case) "Which photo did I plan to share?"
 [9] (drawer) "What snack did I keep for visitors?"
 [10] (post-it) "What conversation starter did I save?"
 [11] (keyboard) "Which emoji set is our team in-joke?"
 [12] (mouse) "What game night invite did I draft?"
 [13] (calendar) "Which lunch roulette slot is mine?"
 [14] (desk surface) "What open-door hours did I set?"
 [15] (monitor bezel) "Who is my co-working buddy?"

Office Breakroom or Pantry

[1] (espresso machine) "Who did I chat with here?"
 [2] (French press) "What brew-and-bond tradition did I start?"
 [3] (menu board) "What drink is the team favorite?"
 [4] (open hours) "When is our coffee club?"
 [5] (paper cups) "Whose name did I write on a cup?"
 [6] (plant) "What photo did we take by the plant?"
 [7] (fridge) "What shared treat did I bring?"
 [8] (water pitcher) "Which hallway hello route did I pair with refills?"
 [9] (logo sign) "What selfie spot did we choose?"
 [10] (kettle) "What tea tasting did we plan?"
 [11] (countertop) "What potluck sign-up did I make?"
 [12] (espresso cups) "What tiny-spoon toast did we say?"
 [13] (bathroom sign) "What 'walk-and-talk' route did we use?"
 [14] (coffee prices) "What budget round did I buy?"
 [15] (grinder) "Who taught me to grind beans?"

Cafe or Bistro

[1] (menu poster) "Who recommended this dish?"
 [2] (garlic sauce) "Who likes this sauce most?"
 [3] (wasabi soy) "What joke did we make about wasabi?"
 [4] (bowls) "Who did I share noodles with?"
 [5] (water bottle) "Who always asks for extra water?"
 [6] (cork jar) "What event did we celebrate?"
 [7] (fork \& spoon) "What utensil game did we play?"
 [8] (glasses) "What toast line did we use?"
 [9] (napkins) "What doodle did a friend draw?"
 [10] (tablecloth grid) "What seating pattern worked best?"
 [11] (Asian Food logo) "What photo tag did we use?"
 [12] (bathroom sign) "What BRB signal did we agree on?"
 [13] (plant) "Who posed with the table plant?"
 [14] (right seat) "Who usually sits on the right?"
 [15] (left seat) "What first-timer topic did I save?"

2.3 Persona 3: Health-conscious Professional

Persona/Role: Healthy enjoyment of life is the highest priority. Every move is all about calories, nutrition, cooking recipes, and health. The well-being of life matters a lot to you.

Seeded Questions per Location:

Format: [index] (object-of-inquiry) "question"

Office Desk

[1] (water bottle) "How many refills did I plan today?"
 [2] (calendar) "What workout time did I set?"
 [3] (desk lamp) "What eye-break rule did I save?"
 [4] (mug) "How many calories did I log for sweet coffee?"
 [5] (keyboard) "What stretch interval did I choose?"
 [6] (mouse) "What wrist-break reminder did I set?"
 [7] (monitors) "What brightness mode did I prefer?"
 [8] (lanyards) "What step goal did I set for event days?"
 [9] (glasses case) "What blue-light cutoff time did I note?"
 [10] (drawer) "What healthy snack did I store?"
 [11] (post-it) "What breathing routine did I write?"
 [12] (right monitor) "What posture cue did I anchor here?"
 [13] (keyboard) "What caffeine cutoff did I set?"
 [14] (calendar) "What 'no-late-email' boundary did I log?"
 [15] (desk surface) "What clean-desk rule reduces snacking?"

Office Breakroom or Pantry

[1] (espresso machine) "What max shots per day did I allow?"
 [2] (French press) "What brew size did I pick?"
 [3] (menu board) "Which drink had the lowest calories?"
 [4] (fridge) "What snack did I plan to take?"
 [5] (kettle) "Which herbal tea did I choose?"
 [6] (filter pitcher) "How often should I refill water?"
 [7] (paper cups) "What sugar rule did I set?"
 [8] (plant) "What pause-and-breathe cue did I save?"
 [9] (coffee prices) "What milk swap did I choose?"
 [10] (bathroom sign) "What walk route did I plan?"
 [11] (countertop) "What 'eat seated only' rule did I log?"
 [12] (sweeteners) "Which sweetener did I prefer?"
 [13] (logo sign) "Which pastry did I mark 'avoid'?"
 [14] (open hours) "What light-snack window did I choose?"
 [15] (espresso cups) "What 'no sugar after' time did I set?"

Cafe or Bistro

[1] (menu poster) "Which dish was my lowest-calorie pick?"
 [2] (garlic sauce) "What serving size did I allow?"
 [3] (wasabi soy) "What sodium note did I record?"
 [4] (bowl) "What portion size did I choose?"
 [5] (water bottle) "How many sips before ordering?"
 [6] (fork \& spoon) "What mindful-eating tip did I use?"
 [7] (napkins) "What slow-down cue did I attach?"
 [8] (glasses) "What 'no-sugary-drinks' rule did I set?"
 [9] (tablecloth grid) "What 'half-plate veg' rule did I log?"
 [10] (Asian Food logo) "Which fried item is 'occasional only'?"
 [11] (plant) "What stress-reduction breath did I save?"
 [12] (bathroom sign) "What post-meal walk did I plan?"
 [13] (right seat) "What 'share a dish' plan did I set?"
 [14] (left seat) "What 'stop at 80% full' cue did I note?"
 [15] (cork jar) "What 'no alcohol at lunch' rule did I log?"

REFERENCES

- [1] Google. GoogleAI. Gemini models. <https://ai.google.dev/gemini-api/docs/models/>. Mar. 21. 2025. 1
- [2] J. Lee, J. Wang, E. Brown, L. Chu, S. S. Rodriguez, and J. E. Froehlich. GazePointAR: a context-aware multimodal voice assistant for pronoun

disambiguation in wearable augmented reality. In *Proc. of ACM CHI*, pp. 1–20, 2024. 1

[3] Y. A. Malkov and D. A. Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE TPAMI*, 42(4):824–836, 2018. 1

[4] Nominatim. Osm-search. <https://github.com/osm-search/Nominatim>, 2025. Aug. 29. 2025. 1

[5] OpenAI. OpenAI. Gpt-4v(ision) system card. https://cdn.openai.com/papers/GPTV_System_Card.pdf. Mar. 28. 2024. 1