AROMAGEN: Interactive Generation of Rich Olfactory Experiences with Multimodal Language Models

Anonymous Author(s)

Abstract

Smell's deep connection with food, memory, and social experience has long motivated researchers to bring olfaction into interactive systems. Yet most olfactory interfaces remain limited to fixed scent cartridges and pre-defined generation patterns, and the scarcity of large-scale olfactory datasets has further constrained AI-based approaches. We present AROMAGEN*, an AI-powered wearable interface capable of real-time, general-purpose aroma generation from free-form text or visual inputs.* AROMAGEN *is powered by a multimodal LLM that leverages latent olfactory knowledge to map semantic inputs to structured mixtures of 12 carefully selected base odorants, released through a neck-worn dispenser. Users can iteratively refine generated aromas through natural language feedback via in-context learning. Through a controlled user study ($N = 26$),* AROMAGEN *matches human-composed mixtures in zero-shot generation and significantly surpasses them after iterative refinement, achieving a median similarity of 8/10 to real food aromas and reducing perceived artificiality to levels comparable to real food.* AROMAGEN *is a step towards real-world interactive aroma generation, opening new possibilities for communication, wellbeing, and immersive technologies.*

1. Introduction

Smell is deeply entangled with the human experience. It shapes how we taste food [7], recall memories [20], and even form social bonds [13]. These qualities have long motivated Human-Computer Interaction (HCI) researchers and artists to bring smell into interaction design, from immersive VR [5, 42] and gaming systems [8, 43] to health [46] and wellbeing [3]. Despite decades of work, olfactory interfaces remain constrained to fixed scent cartridges and pre-defined generation patterns, without the flexibility and fidelity to generate everyday aromas. Existing devices typically rely on a fixed set of preloaded cartridges or reservoirs [9, 21]. Unlike vision or audition, smell has no identified "primaries," meaning designers and engineers cannot flexibly compose or reproduce arbitrary aromas. Compounding this, the scarcity of large-scale olfactory datasets has historically limited the development of AI models for aroma generation. The result is that current systems can only deliver a small, predetermined set of olfactory cues, limiting their scope and applicability.

To address this gap, we present AROMAGEN, an intelligent system powered by a large multimodal model capable of mapping food substances into a distribution of 12 base odorants[1]. Users can describe a target aroma through text (e.g., "*a smoky cheeseburger with caramelized onions*"), upload an image (e.g., *a photograph of a fresh garden salad*), or speak aloud (e.g., "*this chai latte smells spiced and creamy*"), and AROMAGEN leverages the internal knowledge and representations of large language and multimodal models to translate these inputs into actionable aroma blends. The 12 base odorants are carefully selected through formative studies with expert olfactory researchers to ensure sufficient diversity and maximize complementarity of their mixtures. Through a neck-worn wearable dispenser, these selected base odorants can be stored, sequenced, and released as mixtures in real time. We further develop a human-in-the-loop feedback mechanism that enables interactive aroma adjustments through natural language commands (e.g., "*make it stronger*", "*less sweet*", "*more fresh*"), implemented via in-context learning from a few rounds of user feedback. Together, our approach combines multimodal semantic inputs, pre-trained foundation models, and wearable actuation into a closed-loop pipeline for real-time aroma generation.

Through our user studies, we found that this 'language' for aroma enables real-time and general-purpose aroma generation across diverse food stimuli. In a controlled human-subject study ($N = 26$), participants judged reproduced aromas against their real counterparts, with AROMAGEN achieving a median similarity score of 8.0/10 after iterative refinement, significantly outperforming both human-composed aroma blends (Mdn $= 5.5$) and zero-shot generation (Mdn $= 6.0$), with large effect sizes in both cases ($r = .886$).

To summarize, our contributions are as follows:

- A novel aroma generation approach that leverages latent

[1] In olfactory science, an *odorant* is any gas (volatile molecule or combination of molecules) that can be perceived as an aroma when inhaled or exhaled.

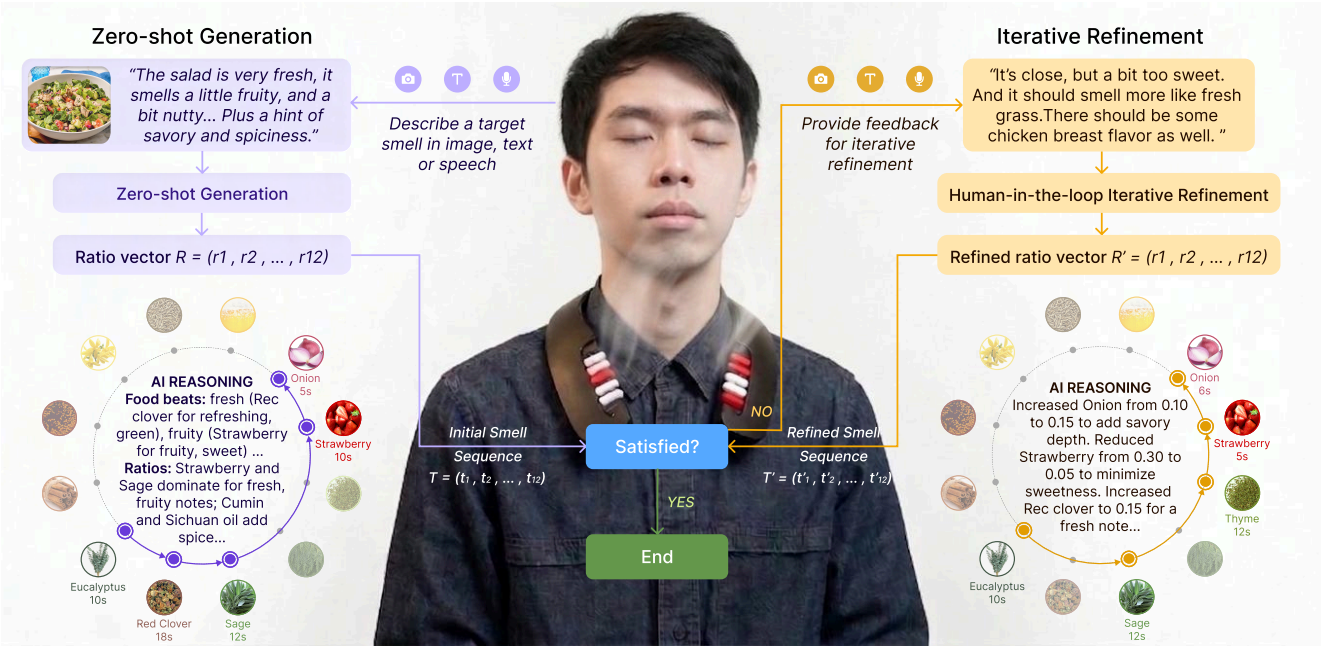

Figure 1. AROMAGEN is an AI-powered wearable interface for real-time, general-purpose aroma generation from free-form text, image, or speech inputs. Given a user description (e.g., "the salad is very fresh, it smells a little fruity and savory"), AROMAGEN performs zero-shot generation to generate a ratio vector over 12 base odorants, which are sequentially released through a neck-worn dispenser. If unsatisfied, users provide natural language feedback (e.g., "a bit too sweet, more fresh grass, some chicken flavor"), and AROMAGEN revises the composition via iterative refinement. This closed-loop pipeline enables intuitive, semantically driven aroma generation without any sensor measurements or chemical specifications.

olfactory knowledge in multimodal LLMs to map free-form text and visual inputs to structured odorant mixtures, requiring no sensor measurements or chemical specifications.

- A real-time wearable system comprising a multimodal interface and a neck-worn dispenser with 12 carefully selected base odorants, supporting on-demand aroma generation and closed-loop iterative refinement.
- A controlled user study ($N = 26$) validating the effectiveness of AROMAGEN for general-purpose aroma generation.

Together, AROMAGEN is a step towards real-world interactive aroma generation, potentially opening new possibilities for communication, well-being, and immersive technologies.

2. Related Work

Our work builds upon research in multimodal AI, developing machine learning models for olfactory data, systems for aroma reproduction and approximation, and olfactory interfaces.

2.1. Olfaction as an AI Modality

The evolution of large language models from text-only systems to multimodal architectures has substantially expanded the range of modalities that AI can perceive and generate [30]. Today's state-of-the-art multimodal models are capable of jointly processing text, images, video, audio, tactile, physical sensing, physiological data, and other interaction modalities [23, 31]. Despite trained only on language, even language models themselves have shown surprising commonsense knowledge about other modalities and domains, like robotics, social interactions, and physical embodied AI [32]. This has opened up many opportunities for leveraging foundation models in tangible, embodied, and multimodal interaction systems [11]. Smell, however, remains largely absent from the AI landscape.

Olfactory perception begins when airborne molecules bind to receptors in the nose, producing signals the brain interprets as odor. To study this systematically, researchers have assembled datasets that characterize molecules along three layers: chemoinformatic structural representations, physicochemical properties (molecular weight, boiling point, vapor pressure, logP), and human perceptual ratings including semantic descriptors (e.g., *floral*, *fruity*), similarity judgments, and detection thresholds. The Dravnieks Atlas [10] established a foundational 146-descriptor space for monomolecular odorants analyzed with linear methods; Keller and Vosshall [24] extended this with 476 molecules. More recently, Lee et al. [27] represented

Approach	Input Type	Data Type	Algorithm	Output
QCM [38, 51]	Sensor measurement	Oscillation frequency	Linear search, matrix matching	Simple fruit aromas (e.g., apple, orange)
GC-MS [1, 41]	Sensor measurement	MS spectra	Non-negative matrix factorization + Deep neural network	Essential oils, new fragrances
LLM (Ours)	**Natural language / Image**	**LLM latent perceptual knowledge**	**Zero-shot + In-context learning**	**General food aromas**

Table 1. Comparison of aroma reproduction approaches. Prior methods require specialized sensor equipment to measure physical samples; our LLM-based approach instead accepts natural language descriptions or images, requiring no sensors or chemical specifications.

molecules as graphs and trained a message-passing neural network to derive a "Principal Odor Map". Work on aroma mixtures has remained largely descriptive [44, 48], and the scarcity of labeled perceptual data has historically constrained model complexity [14]. Large-scale sensor datasets such as SmellNet [12] open new possibilities for foundation-model approaches to smell sensing.

Our work takes a different approach: rather than predicting perception from molecular features, we leverage the olfactory knowledge already encoded in multimodal LLMs to translate free-form language into actionable aroma formulations.

2.2. Olfactory Language and Semantics

A longstanding assumption in cognitive science holds that smell is uniquely resistant to verbal description, a phenomenon termed *olfactory ineffability* [34]. Studies with English speakers confirm that odor talk is infrequent, dedicated smell terms are rare, and naming familiar aromas is surprisingly difficult even for people with intact olfaction [33]. However, Majid and Burenhult [34] demonstrated that this is not a universal human limitation: speakers of Jahai, a language spoken by hunter-gatherers on the Malay Peninsula, name aromas as easily as colors, suggesting that the smell-language gap is shaped by culture and ecology rather than biology [34].

Separately, Hörberg et al. [22] took a corpus-based approach to map the structure of olfactory language in English, automatically identifying aroma descriptors from natural text and deriving their semantic organization [22]. Despite the sparse aroma vocabulary of English, they found a coherent semantic space structured primarily along dimensions of pleasantness and edibility, suggesting that latent olfactory knowledge is nonetheless encoded in natural language. This view is further supported by recent work showing that large language models, particularly GPT-4o, can recover olfactory-semantic relationships from text to a degree that correlates with human perceptual judgments [26].

Together, these findings motivate our approach: rather than requiring sensor measurements or molecular specifications, we leverage the olfactory knowledge already latent in multimodal LLMs to bridge natural language and smell.

2.3. Aroma Reproduction and Approximation

Aroma reproduction research investigates how to recreate a target aroma by blending known components so that the resulting mixture is perceptually similar to the original. Unlike vision or audition, olfaction has no set of primaries from which arbitrary percepts can be synthesized [15]; reproduction therefore relies on searching a mixture space over a finite component palette.

Nakamoto [38] introduced the concept of an *odor recorder* that uses QCM sensor arrays to capture a fruit odor's signature and then searches mixture ratios over a small component set to approximate it [37, 38, 51]. For example, orange (originally produced from 14 ingredients) could be convincingly approximated with only three (linalool, citral, and decanal) [52]. Complex essential oils have similarly been approximated using 10–30 components with non-negative matrix factorization and mass spectrometry [39, 41]. More recent algorithmic frameworks combine odorant descriptor prediction with optimization: Nicolaï et al. [40] demonstrated real-time aroma reconstruction of citrus varieties using 12 components, and deep learning approaches combine mass spectra with descriptor prediction and gradient descent to design new formulations [1].

A critical limitation shared by all these approaches is that they require either sensor measurements of a physical target sample or precise molecular specifications as input: none accept the open-ended natural-language descriptions that people use in everyday life to communicate about smell.

2.4. Olfactory Interfaces

Work on olfaction in HCI has primarily focused on *output*: designing devices that release aromas to enrich interactive experiences. Motivated by smell's strong links to memory, emotion, and presence [17, 19], researchers have explored applications spanning healthcare, entertainment, and social interaction—including breathing guidance [54], motion sickness reduction [45, 47], narrative enrichment [8], social collaboration [28, 36], and body perception alteration [4]. Delivery mechanisms range from thermal diffusion and aerosolization [2, 42, 53] to directed airflow, ultrasound-steered vortices [18], and body-worn devices including on-face wearables [50], trigeminal nose clips [6], and retronasal aroma displays for eating and drinking [29, 35].

Despite this breadth, virtually all olfactory interfaces rely on a fixed repertoire of pre-selected aromas, precluding dynamic, user-specified aroma generation at runtime. Our system removes this constraint by coupling an LLM-based formulation pipeline with an existing multi-channel olfactory display, enabling users to request any food aroma in

natural language and receive a synthesized approximation on demand.

3. Formative Studies

To ground key design decisions in human olfaction perception and practice, we conducted two formative studies. Study 1 examined how non-experts describe and decompose aromas in language, and whether commercial essential oils are perceptually adequate as base odorants. Building on these findings, we deployed an early prototype at a hackathon, in which the system mapped multimodal user inputs (text and images) to mixtures of base odorants via a multimodal LLM. Study 2 then conducted expert interviews with domain specialists to validate and refine concrete design decisions around palette selection, aroma sequencing, and on-device reproduction.

3.1. Study 1: Olfactory Perception and Language

To understand how non-experts describe and decompose aromas, and whether natural and synthetic odorant materials are perceptually adequate as base odorants, we conducted a lab study with 10 participants using four common food stimuli.

3.1.1. Protocol

We recruited 10 participants (P1–P10, 3 female, 7 male, ages 20–40) with no professional background in perfumery or chemistry, compensated with a $10 gift card. All confirmed the absence of anosmia or relevant allergies. Sessions lasted 20–30 minutes and comprised three tasks followed by a brief semi-structured interview. The study was IRB-approved and all sessions were audio-recorded and transcribed with consent.

Participants completed the following tasks using four photographs of common aroma-bearing objects (strawberry, coffee beans, burger, banana), selected to span simple to complex aroma profiles:

Task 1 — Aroma Description: Participants described each aroma in their own words, thinking aloud.

Task 2 — Odorant Composition: Participants decomposed each aroma into constituent components and allocated 7 points across them to reflect perceived contribution.

Task 3 — Synthetic Fidelity: Participants smelled each of four aromas, including three natural essential oils (coffee, strawberry, and banana) and one synthetic mixture (burger, composed of isovaleric acid and sweet orange), alongside their real food counterpart and rated perceptual similarity on a 7-point scale.

Two researchers independently coded transcripts using an inductive approach [49], supplemented by word frequency analysis (>2 occurrences) to surface shared perceptual vocabulary.

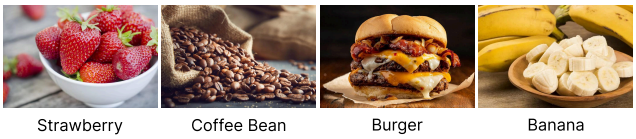

Figure 2. Formative study 1 stimuli.

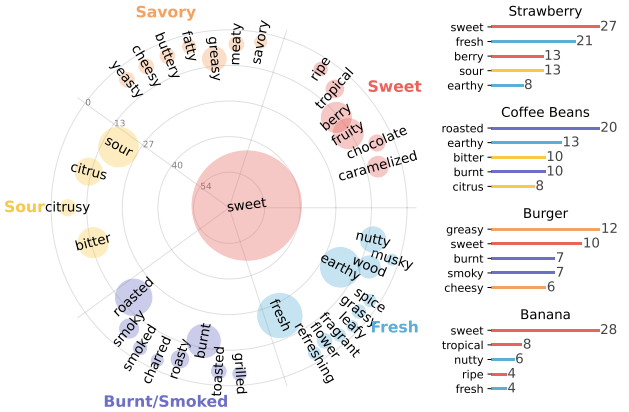

Figure 3. Polar chart of aroma descriptors elicited across Formative Study 1, grouped by olfactory category. Word proximity to center reflect frequency of use across all 10 participants. Recurring descriptors such as *sweet*, *roasted*, and *fresh* suggest a shared perceptual vocabulary that informed the selection of base odorants in AROMAGEN.

3.1.2. Findings

Aroma description relies on borrowed vocabulary and ingredient-level reasoning. Participants lacked dedicated olfactory vocabulary, borrowing from taste, texture, and memory (e.g., P1 described coffee as "*leathery*" and "*dry*"). For composite foods, participants reasoned about visible ingredients (P9 listed onion, meat, bacon, and cheese for cheeseburger), but could not decompose single-substance stimuli (P3, P4 described banana as "*fundamental*" and simply "*smells like banana*"). This positions LLMs, trained on rich cross-modal language, as well-suited to translate imprecise descriptions into structured aroma representations.

A shared perceptual vocabulary underlies aroma description. Word frequency analysis across participants and stimuli yielded 36 unique terms, with *sweet* dominating at 68 occurrences, nearly three times the next most frequent term *fresh* (Figure 3), suggesting a constrained, convergent olfactory vocabulary. As sentence-transformer embeddings failed to recover perceptually meaningful olfactory groupings during PCA with K-Means clustering, two researchers manually grouped the 36 terms through thematic analysis, converging on five perceptual categories: *sweet*, *savory*, *sour*, *burnt/smoked*, and *fresh*. These categories directly informed our selection of base odorants for AROMAGEN's palette.

Essential oils approximate real aromas directionally. Banana and strawberry were most recognizable (avg. ∼4.8/7

ID	Background	Relevant Experience
E1	Perfumer	Scent design for film screenings and exhibitions; trained in the IFF–ISIPCA MSc in Scent Design and Creation.
E2	Fragrance Formulator	7 years in candle layering and body perfumery; specializes in layering natural and synthetic molecules.
E3	Perfumery Educator	Operates a DIY perfumery workshop; worked with ∼500 clients over 2 years; background in computer science.

Table 2. Demographics of fragrance experts interviewed in Study 2.

and ∼3.9/7). However, all essential oils were perceived as overly sweet and artificial, likened to candy or medicine (P02, P03, P04). The burger (isovaleric acid + sweet orange) scored the lowest (avg. ∼1.8/7), described as "Orbeez" with no meat character (P09). This validates essential oils as directionally adequate base odorants while motivating the iterative refinement mechanism in AROMAGEN.

3.2. Early Prototype

Building on the findings of Study 1, we deployed an early prototype at a hackathon ($N \approx 20$) to probe the feasibility of LLM-driven aroma generation in a naturalistic setting. The prototype used a fixed palette of six heuristically chosen base odorants, mapping natural language descriptions to a sequential diffusion cycle. Prompting strategies and per-odorant dispensing durations were iteratively refined through user feedback. While participants responded positively to abstract prompts, perceived fidelity broke down when descriptions referenced specific food items, motivating a need to validate our design decisions against domain expertise.

3.3. Study 2: Expert Interview

Building on Study 1 and the early prototype, we interviewed fragrance experts to inform our design decisions.

3.3.1. Protocol

We recruited three fragrance experts (Table 2) . Each participated in a 20–30 minute semi-structured interview covering four topics: (1) the system's base odorant selection and (2) on-device reproduction strategy. Interview questions are provided in Appendix A.

3.3.2. Findings

Synthetic and natural odorants are adequate base materials, and a compact palette suffices within a defined domain. Experts confirmed that both natural extracts and synthetic chemicals are standard media for consumer-facing blending applications and remain perceptually adequate within hardware and cost constraints (E1, E2, E3). On palette size, while professional perfumers often work with thousands of materials, experts agreed that a well-chosen compact set anchored to a specific domain yields meaningful coverage (E1, E2); E3 noted that 12 base odorants appeared reasonable for food aroma reproduction.

Volatility-ordered sequential delivery and limiting active odorants are key to on-device reproduction. Experts described volatility as the primary driver of temporal aroma perception: lighter notes should be released first, as opening with a heavy note renders subsequent lighter notes imperceptible due to olfactory adaptation (E2). E3 confirmed that for food environments, releasing stronger or spicier notes first aligns with how aromas are naturally perceived. On exposure duration, experts converged on one to two minutes as the appropriate window, as shorter exposures are insufficient for initial perception while longer ones lead to olfactory adaptation (E1, E2). For ingredient count, experts consistently recommended limiting active odorants per blend to 3–6, enough to capture layered food character without inducing perceptual overload (E1, E2, E3).

3.4. Design Implications

The formative studies informed three key design decisions in our AROMAGEN system.

[D1] LLM-based semantic-to-odorant mapping. Participants described aromas at the semantic level rather than at the component level (Study 1). Since LLMs encode rich cross-modal associations between language and sensory experience, we can use a multimodal LLM to map high-level descriptions directly to mixtures of base odorants, bypassing the need for explicit chemical specifications.

[D2] Base odorants and generation. Expert interviews (Study 2) confirmed that a small set of odorants guided by perceptual dimensions (Study 1) provides adequate coverage within the food domain, that 3–6 active ingredients per blend represent the perceptual sweet spot, and that a 60-second dispensing cycle with volatility-ordered sequential release is sufficient for on-device aroma reproduction.

[D3] Iterative refinement. We introduce a closed-loop feedback mechanism in which users iteratively refine the generated blend through natural language, compensating for the inherent limitations identified in Study 1.

4. AROMAGEN

AROMAGEN enables on-demand generation of realistic aromas for any food, leveraging the broad olfactory knowledge encoded in a multimodal LLM. Users describe a target aroma via text, image, or speech; the LLM maps the input to a ratio vector over 12 base odorants using zero-shot generation, which is dispensed through a wearable device. Users then provide natural language feedback to iteratively refine the

Figure 4. The 12 base odorants used in AROMAGEN's palette.

Odorant	Vol.	Notes
Cumin	6	Smoky, spice
Ylang Ylang	6	Warm, light spice
Sichuan Oil	3	Light, chai, spice
Cinnamon	5	Sweet, spice, coffee, warm
Eucalyptus	5	Refreshing, spa
Red Clover	5	Mint, clover, green, refreshing
Sage	6	Refreshing
Cypress	5	Woody stability
Thyme	5	Bitter, green, vegetable
Strawberry	5	Elegant clarity, fruity, sweet
Onion	6	Umami, onion, chips, savory
Isovaleric Acid	8	Cheesy, sweat, sour

Table 3. The 12 base odorants in AROMAGEN's palette, with volatility score (1–10) and characteristic notes. Colored squares indicate perceptual categories: Sweet, Savory, Sour, Burnt/Smoked, Fresh.

mixture, forming a closed-loop pipeline that generalizes across food types without requiring pre-programmed aroma libraries.

4.1. Aroma Space Design

We selected 12 base odorants guided by the five perceptual categories surfaced in Study 1 and prior work on aroma mixture approximation [38, 40]. The number 12 reflects the maximum capacity of our wearable hardware, and selection followed three heuristics: (1) *perceptual coverage*, minimizing overlap across aroma space; (2) *mixability*, favoring materials that combine without aversive off-notes; and (3) *volatility balance*, spanning volatility levels from 1 to 10 to produce naturalistic temporal dynamics. The palette comprises both natural materials (essential oil, tincture, fragrance oil, and powder) and synthetic aroma chemical, selected to maximize perceptual diversity within hardware and cost constraints (see Appendix C for ingredients and Appendix E for material type definitions).

Each odorant is encoded in a structured JSON file containing its name, volatility score, semantic notes, and device channel location, which is provided to the LLM as reference during aroma composition.

4.2. AI Modeling of Aroma Composition

Given the aroma space, we develop an AI-based approach to automatically infer its compositions from user inputs. We start with a pure zero-shot generation approach, leveraging domain knowledge encoded in large language models for aroma mixture inference. We then extend this to iterative refinement, incorporating a few rounds of user feedback to personalize and refine the generated compositions.

4.2.1. Zero-shot Generation

Our first approach is purely zero-shot, in which no labeled demonstrations of odorant breakdowns are provided. Instead, the model relies entirely on task instructions specified in the system prompt and uses its internal knowledge to infer odorant mixtures given a context.

We formulate aroma composition as a structured prediction task, where the model maps a multimodal user description to a ratio over d base odorants. Formally, given a user input x consisting of a text description, an optional image processed via a cascaded image-to-text pipeline, and optional transcribed speech, the model produces a ratio vector $\mathbf{r} = (r_1, \ldots, r_d)$ where $r_i \geq 0$ and $\sum_{i=1}^{d} r_i = 1$. Each ratio r_i is then mapped to a dispensing duration $\tau_i = r_i \cdot T$, where $T = 60$ seconds is the total dispensing cycle, such that $\sum_{i=1}^{d} \tau_i = T$.

We prompt the LLM using a fixed two-component template: (1) a system instruction defining the task and output format (see Appendix B), and (2) the user query. The system instruction provides the full odorant palette with qualitative attributes including volatility and perceptual character, and instructs the model to decompose any described aroma into a convex combination of $d = 12$ base odorants, returned as a structured JSON object summing to 1. The model is directed to first identify the dominant aroma of the input via the `note` field, extract perceptual beats (e.g., food category, preparation state, key notes), and allocate ratios by prioritizing high-volatility odorants for primary recognition, using 3–6 active odorants and setting the remainder to 0.

4.2.2. Iterative Refinement

Zero-shot generation is lightweight but risks misinterpreting user context or failing to capture individual preferences. We therefore extend zero-shot generation to integrate learning from users: after each dispensing cycle, the user evaluates the resulting aroma and provides natural language feedback to refine the composition. The system treats each refinement as an incremental update conditioned on the full interaction history, placing the human in the loop to address the inherent subjectivity of olfactory experience, which cannot be fully captured by any fixed model.

We implement iterative refinement via in-context learning (ICL), allowing the LLM to adapt to user preferences at

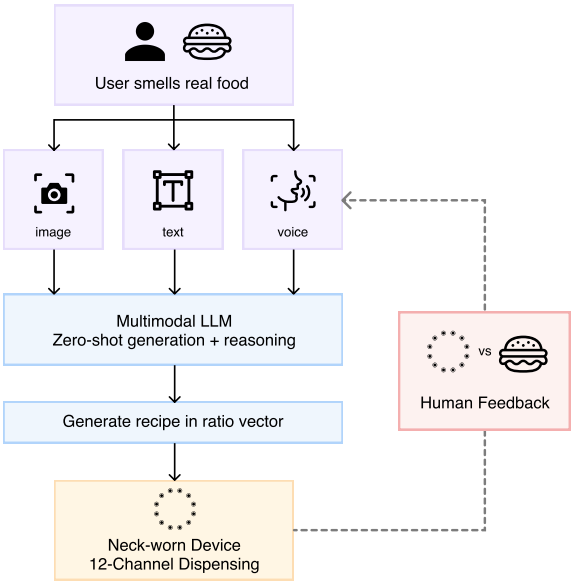

Figure 5. The AROMAGEN system pipeline. Users initiate zero-shot generation via multimodal inputs (text, image, or speech), which the system translates into an initial odorant mixture vector. Through human-in-the-loop iterative refinement, users can refine the aroma using natural language.

inference time by conditioning on accumulated interaction history without gradient updates. Formally, at iteration t, the LLM receives a structured prompt containing the original food description, the current ratio vector $\mathbf{r}^{(t-1)}$, the full history of prior feedback and corresponding changes $\{(\text{feedback}^{(i)}, \text{changes}^{(i)})\}_{i=1}^{t-1}$, and the latest user feedback $\text{feedback}^{(t)}$. The model is constrained to perform targeted revisions: preserving ratios the user did not criticize, adjusting only what the feedback demands, and preferring shifts in existing odorant ratios over introducing new ones.

The session data logged at each iteration, including modalities used, ratio vectors, feedback text, and response times, provides a foundation for future personalization. Over repeated sessions, this interaction history could be leveraged to adapt the system to individual user preferences, reducing the number of calibration steps required to reach a satisfactory composition. An example of the iterative refinement process is shown in Figure 6.

4.3. Interaction Design

4.3.1. Interface Design

AROMAGEN is implemented as a web-based interface (Figure 7) organized into a circular orbital display and an input panel. Users submit descriptions via text, voice (auto-transcribed via Whisper), or image (auto-described via GPT-5), combinable within a single submission. Upon generation, active odorants appear as labeled nodes around the orbital ring annotated with dispensing durations. Users refine the result by providing feedback via the same panel, routed to

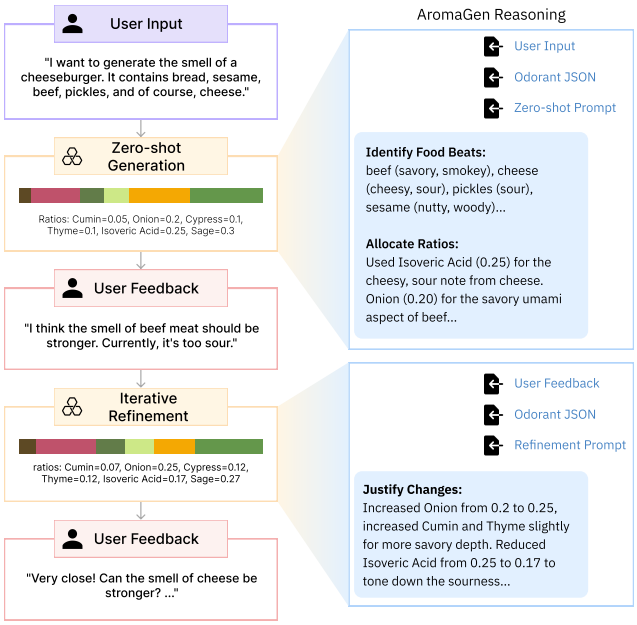

Figure 6. An example of AROMAGEN's iterative refinement and internal reasoning process: semantic decomposition (e.g., identifying food components), projection into a perceptual space (e.g., savory, sour), and constrained allocation to a ratio vector over base odorants. User feedback is incorporated via in-context learning, where high-level adjustments (e.g., "less sour") are translated into targeted updates of the aroma mixture.

the in-context learning pipeline. A *Play on Device* button dispatches the composition to the hardware controller, initiating a 60-second dispensing cycle.

4.3.2. Hardware Implementation

AROMAGEN is built upon the SCENTAC X-Scent 3.0, a neck-worn digital scent player housing 12 channels (DC 5V, <380 g). We replaced the original cartridges with custom modules filled with our base odorants. Each channel is independently driven by a fan diffuser that directs airflow through an odorant-soaked medium, releasing aromas for a duration proportional to its predicted ratio from the AI backend, which dispatches channel-specific BLE commands sequentially to the device. Odorants are released in volatility-descending order, consistent with expert recommendations from Study 2.

5. Experiments

We evaluate our proposed AROMAGEN system through both quantitative evaluation and user studies.

5.1. User Study

To evaluate AROMAGEN, we conducted a controlled user study addressing three research questions:

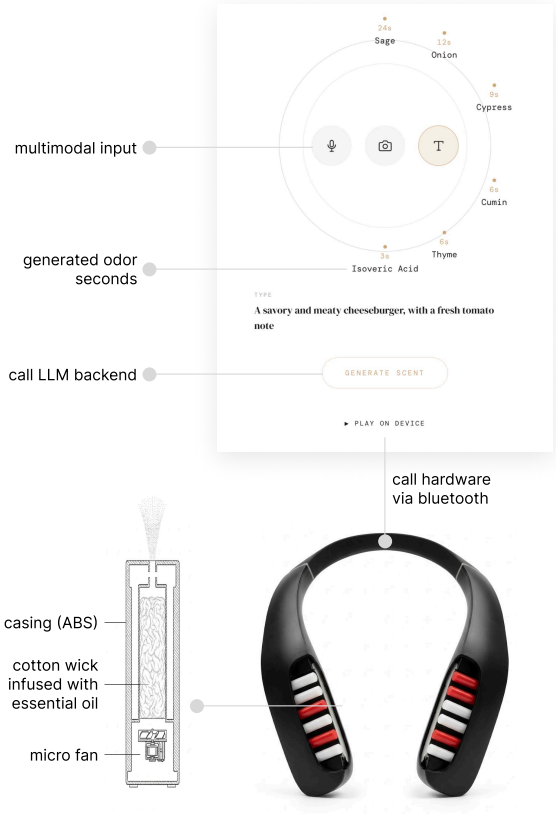

Figure 7. We developed a lightweight web interface enabling users to submit text descriptions and images of food substances, inspect the generated aroma composition, play it on the hardware device, and iteratively refine through feedback.

RQ1: Does AROMAGEN produce mixtures that are perceptually closer to real food than human-only mixing?

RQ1.1: Do zero-shot generation and iterative refinement each improve upon the human-only mixing baseline?

RQ1.2: Does iterative refinement further improve upon zero-shot generation?

RQ2: Does any specific olfactory dimension drive perceptual differences across conditions?

RQ3: Does AROMAGEN reduce cognitive load compared to human-only mixing?

5.2. Methods

5.2.1. Participants

We recruited 26 participants (10 male, 16 female, ages 19–49, $M = 27.1$, $SD = 6.6$) via university email and flyers. All self-reported normal olfactory function, no relevant allergies, and no professional background in perfumery or flavoring. Participants received a $15 gift card and a perfume sample.

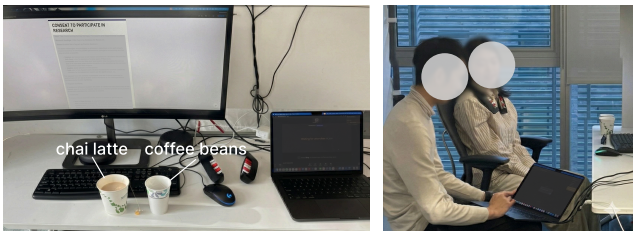

Figure 8. User study setup. Coffee beans are provided for participants to get refreshed in the process.

Figure 9. User Study Stimuli

5.2.2. Stimuli

We selected three foods spanning diverse aroma profiles: **Pizza** (roasted, savory), **Salad** (fresh, mixed), and **Chai Latte** (spiced, dairy), shown in Figure 9. All items were freshly prepared on the day of each session; Pizza and Chai Latte were microwaved for 30 seconds immediately before each trial to ensure consistent aroma diffusion.

5.2.3. Conditions

Each participant completed four conditions per food stimulus:

- ***Real***: Participants smelled the actual food, serving as perceptual ground truth.
- ***Human***: Participants manually composed an aroma mixture by estimating compound ratios from direct reference to the real food, without AI assistance.
- ***w/o Learning***: The AROMAGEN system generated compound ratios from participant multimodal (text, image) input via zero-shot generation, with no iterative refinement.
- ***AromaGen***: The AROMAGEN system generated compound ratios from participant multimodal (text, image) input via zero-shot generation, then participants iteratively refined the composition via human-in-the-loop feedback, with each round conditioning on the full prior interaction history through in-context learning.

5.2.4. Procedure

After a brief tutorial on the AROMAGEN interface and available compounds, participants completed all four conditions in fixed order (*Real* → *Human* → *w/o Learning*→ *AromaGen*), thinking aloud throughout and providing measure-

	Similarity Scores (↑)	Semantic Dist. to *Real* (↓)
Human Mdn (IQR)	5.5 [4.0, 7.0]	5.39 [3.32, 7.75]
w/o Learning Mdn (IQR)	6.0 [5.0, 7.0]	5.20 [4.00, 7.68]
AromaGen Mdn (IQR)	8.0 [6.2, 8.8]	3.32 [2.24, 6.40]
Friedman	$\chi^2 = 25.89, p < .001$	$\chi^2 = 9.61, p = .008$
Human vs *w/o Learning*	$p = .098$, ns	$p = .886$, ns
w/o Learning vs *AromaGen*	$p = .001, r = .886^{**}$	$p = .001, r = .665^{**}$
Human vs *AromaGen*	$p < .001, r = .886^{***}$	$p = .002, r = .662^{**}$

Table 4. Similarity scores and semantic descriptor distance to *Real* by condition. *AromaGen* significantly outperformed both *Human* and *w/o Learning* on both measures (all $p \leq .002$), achieving a median similarity of 8.0 and the lowest semantic distance to *Real*, demonstrating that iterative refinement substantially improves perceptual fidelity to real food aroma. $^{**}p < .01$, $^{***}p < .001$.

ments after each condition. A semi-structured interview followed upon completion. Sessions were audio-recorded with consent and lasted approximately 30 minutes.

5.2.5. Measurements

- **Semantic descriptors:** Building on the five perceptual categories identified in Study 1, we added a sixth dimension *chemical/artificial* to capture synthetic artifacts, yielding six olfactory dimensions (*sweet*, *savory*, *sour*, *burnt/smoked*, *fresh*, *chemical/artificial*) rated on a 10-point Likert scale, validated against the Dravnieks Atlas [10] and DREAM Challenge [25].
- **Similarity score:** After smelling *Real* as reference, participants rated the perceptual similarity of *Human*, *w/o Learning*, and *AromaGen* to the real food on a 10-point Likert scale.
- **NASA-TLX:** Cognitive load assessed after *Human* and *AromaGen* conditions across six subscales [16].
- **Interaction logs:** User inputs and full feedback sequences in *w/o Learning* and *AromaGen*, including iteration count and refinement content.
- **Interviews:** Semi-structured interviews conducted after all conditions.

5.2.6. Analysis

Ordinal measures were analyzed using Friedman tests with post-hoc pairwise Wilcoxon signed-rank tests (FDR-corrected). NASA-TLX used Bonferroni correction ($\alpha = 0.05/7 \approx .007$). For semantic descriptors, we computed Euclidean distance from each condition to *Real* in the six-dimensional descriptor space as a measure of perceptual divergence. Effect sizes are reported as $r = |Z|/\sqrt{N}$.

5.3. Quantitative Results

5.3.1. AI Improvement (RQ1)

Zero-shot generation (RQ1.1). *w/o Learning* achieved similarity ratings comparable to *Human* (Mdn $= 6.0$ vs. 5.5; $p_{\text{fdr}} = .098$, ns), with no statistically significant difference

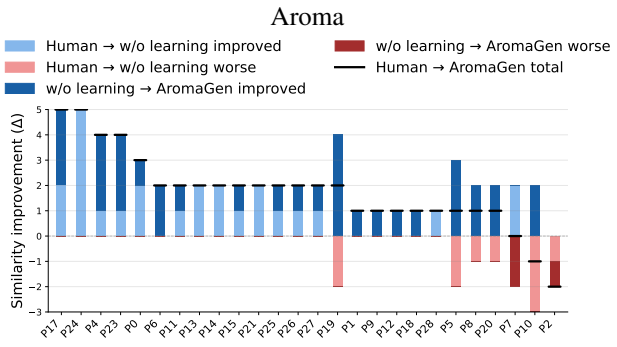

Figure 10. Per-participant similarity improvement across conditions. Bars show incremental changes across *Human*→AROMAGEN *w/o learning*→AROMAGEN; horizontal lines indicate overall *Human*→AROMAGEN gain. The majority of participants improved monotonically, demonstrating that iterative refinement in AROMAGEN consistently drives similarity closer to the real aroma.

between the two conditions. The semantic descriptor analysis corroborated this pattern: Euclidean distance to *Real* did not differ significantly between *w/o Learning* and *Human* (Mdn $= 5.20$ vs. 5.39; $p_{\text{fdr}} = .886$, ns). These results suggest that despite never being explicitly trained on olfactory composition tasks, the LLM encodes sufficient cross-modal knowledge to match non-expert human performance, requiring no sensor measurements, chemical specifications, or labeled training data.

Iterative refinement (RQ1.2). Iterative refinement drove the most substantial gains across both measures. *AromaGen* (Mdn $= 8.0$, IQR $= [6.2, 8.8]$) significantly outperformed both *Human* (Mdn $= 5.5$; $p_{\text{fdr}} < .001$, $r = .886$) and *w/o Learning* (Mdn $= 6.0$; $p_{\text{fdr}} = .001$, $r = .886$), with large effect sizes in both cases. Semantic distance corroborated this pattern: *AromaGen* (Mdn $= 3.32$) was significantly closer to *Real* than both *Human* (Mdn $= 5.39$; $p_{\text{fdr}} = .002$, $r = .662$) and *w/o Learning* (Mdn $= 5.20$; $p_{\text{fdr}} = .001$, $r = .665$). These results demonstrate that the model successfully leverages human perceptual feedback through in-context learning: by conditioning on accumulated user judgments at inference time, AROMAGEN progressively refines its aroma formulations toward compositions that more closely match real food aromas, without any gradient updates or retraining.

Similarity progression. At the individual level, most participants showed monotonic improvement across conditions (Figure 10). While six participants declined from *Human* to *w/o Learning*, five of them recovered in *AromaGen*, suggesting that the zero-shot output, even when initially weaker than the human-only mixing baseline, generally provided a workable starting point that iterative refinement could improve upon. Only two participants showed an overall decline from *Human* to *AromaGen*. This pattern supports the robustness of the iterative refinement: users were able to articulate

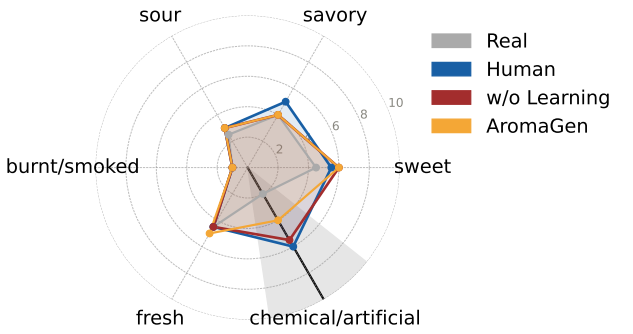

Figure 11. Radar chart of median semantic descriptor ratings across conditions (1–10 scale). *AromaGen* reduced perceived artificiality to a level comparable to *Real* food ($p_{\text{fdr}} = .140$), while *Human* and *w/o Learning* remained significantly more artificial ($p_{\text{fdr}} = .012$ and .015, respectively).

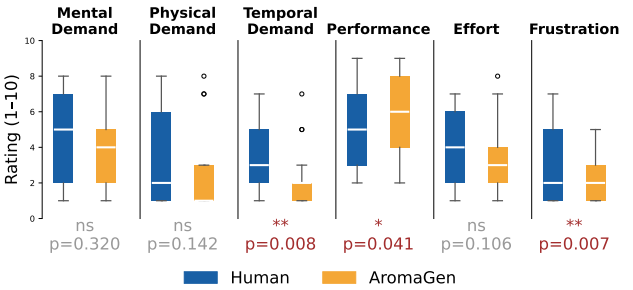

Figure 12. NASA-TLX ratings for *Human* and *AromaGen* conditions. *AromaGen* significantly reduced temporal demand, improved perceived performance, and lowered frustration (all $p < .05$), indicating lower cognitive burden and greater sense of control.

feedback that the model successfully acted upon.

5.3.2. Individual Descriptors (RQ2)

Among the six semantic descriptors, only **chemical/artificial** differed significantly across conditions after FDR correction. Both *Human* and *w/o Learning* were perceived as notably more artificial than *Real* (Mdn = 6.0 and 5.5 vs. 2.0; $p_{\text{fdr}} = .012$ and .015), indicating that neither human-only mixing nor zero-shot generation alone can eliminate the synthetic character of essential oils. In contrast, *AromaGen* (Mdn = 4.0) was not significantly more artificial than *Real* ($p_{\text{fdr}} = .140$), and was rated significantly less artificial than both *Human* ($p = .001$) and *w/o Learning* ($p = .006$). This suggests that iterative refinement is the key factor in reducing perceived artificiality, bringing the aroma profile closer to that of real substances.

5.3.3. Cognitive Load (RQ3)

Since human-only mixing requires users to estimate compound ratios without assistance, we hypothesized that *AromaGen* would reduce cognitive load relative to *Human*. *AromaGen* showed lower temporal demand (Mdn = 2.0 vs. 3.0; $W = 11.0$, $p = .008$), higher perceived performance (Mdn = 6.0 vs. 5.0; $W = 21.0$, $p = .041$), and more consistently low frustration (IQR = 2.0 vs. 4.0; $W = 8.0$, $p = .007$), though none survived Bonferroni correction. Mental demand, physical demand, effort, and composite TLX did not differ significantly (all $p > .05$). Together, these trends suggest that *AromaGen* reduces time pressure and produces more consistent user experiences compared to human-only mixing.

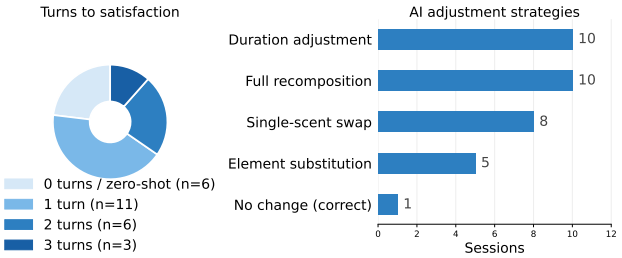

Figure 13. Distribution of refinement turns to satisfaction (left) and AI adjustment strategies observed across feedback sessions (right). Among participants who required refinement ($n = 20$), 85% ($n = 17$) converged within two turns, demonstrating that AROMAGEN consistently produced compositions requiring minimal iterative correction.

5.4. System Evaluation

We analyzed system logs to characterize patterns of how users interacted with AROMAGEN.

Convergence efficiency. Across 26 participants, 88% ($n = 23$) converged within two refinement turns under the *AromaGen* condition, with 23% ($n = 6$) reaching satisfaction from the *w/o Learning* zero-shot generation alone. Among those who provided feedback, the mean number of refinement turns was just 1.60 ($SD = 0.75$, range 1–3), demonstrating that AROMAGEN consistently produced high-quality initial compositions that required minimal correction.

AI reasoning patterns. Analysis of AI reasoning logs in AROMAGEN revealed four recurring adjustment strategies (Figure 13): duration adjustment (10 instances, e.g., extending a savory note to amplify tomato character), full recomposition (10, e.g., rebuilding an entire sequence after a participant found the profile too sweet), single-aroma swap (8, e.g., replacing a herbal note with a woody one to remove an unwanted lemony tone), and element substitution (5, e.g., simultaneously replacing two ingredients to remove an eggy note and rebalance the spice profile). Across these, consistent reasoning behaviors emerged: high-volatility compounds were consistently assigned shorter durations to deliver brief accents without overpowering; and the system replaced disliked ingredients with functionally similar alternatives rather than simply removing them.

5.5. Qualitative Insights

To complement the quantitative results, we analyzed participants' think-aloud protocols and post-study interviews.

AI-generated mixtures are perceived as more layered and unified than human-composed ones (RQ1). Participants found AI compositions formed a coherent whole that exceeded their manual attempts. P26 (chai latte) preferred the AI's sequence because it introduced unexpected but necessary layers: "*I feel like it's more like a black box. So I think it's definitely made some smell that I didn't pick, which made it more similar to the milk tea... I don't try to deconstruct it*". This emergent complexity effectively bridged users' "*vocabulary gap*" in articulating aromas.

Image input captures olfactory nuances that language alone misses (RQ1). Participants frequently found that their own verbal descriptions inadvertently introduced biases that the system faithfully reproduced. P01 (salad) realized their text prompt skewed the generation because it relied on flawed human recall; switching to an image immediately corrected this: "*My text version was way too sweet because I completely missed the acid vinegar note in my description. The image-based one is pretty spot on... The system was better than me at capturing the nuances I missed*".

Iterative refinement enables targeted, semantically driven adjustments that converge quickly (RQ1.2, RQ3). Rather than thinking in terms of specific chemical components, users gave high-level semantic commands like "*less sweet*", "*more grassy*", and found the system responsive. P05 noted that labeled aroma outputs were critical: knowing which component to target enabled precise, confident commands that were impossible during human-only mixing. P10 found it easier to describe the experience than specify parameter changes: "*It is significantly easier to just tell the AI that the fruity note was overwhelming and let it handle the formulation, rather than me trying to manually target and remove the cinnamon*".

Perceived artificiality is the dominant barrier to perceptual realism, and iterative refinement is key to reducing it (RQ2). Participants frequently used terms like "*medicinal*", "*too sharp*", or "*bathroom citrusy*" to characterize generated aromas, indicating an uncanny valley. P15 drew an analogy to fake meat: "*It is exactly like how fake meat tries to be real meat. The general direction is right, but there is this subtle, sharp, almost bathroom-citrusy or perfume-like quality that immediately tells your brain it's artificial*". Participants reported that iterative refinement specifically targeted this artificiality through commands like "*less sharp*" or "*more natural*": "*Once it understood to dial back that artificial base, the whole profile became much more natural*".

Spontaneous ideation reveals generalization potential beyond food scenarios. Unprompted, participants proposed applications such as tracking smell recovery post-COVID as a diagnostic tool (P09, P25), voice-controlled home ambiance diffusion (P20), and restaurant pre-ordering experiences (P05, P08). P26 envisioned "*smell portals*" for sensory transitions in public spaces: "*We have transitions from space to space for almost all other senses, but there's no threshold or portal for smell*". Users also expressed desire to recreate nostalgic environments, such as a favorite beach (P01) or a distant family's kitchen (P13), to bridge physical distances.

6. Discussion

6.1. Toward General-Purpose Aroma Generation

Our results demonstrate that language serves as a viable medium for general-purpose aroma generation: by leveraging the olfactory knowledge already encoded in multimodal LLMs, AROMAGEN translates free-form text and images into actionable aroma mixtures without requiring sensor measurements or chemical specifications. This is a meaningful departure from prior aroma reproduction systems, which are constrained to physical samples or molecular inputs. The success of zero-shot generation across diverse food stimuli, and its further improvement through iterative refinement, suggests that LLMs encode sufficient cross-modal olfactory knowledge to support open-ended aroma generation. Unlike vision or audio, olfactory perception lacks objective ground truth, and what smells "correct" is shaped by individual memory and context. This positions human-in-the-loop iterative refinement not as a workaround for model limitations, but as a structurally necessary component of any general-purpose aroma generation system.

6.2. Usage Scenarios

Beyond the food domain evaluated in our study, AROMAGEN suggests a broader paradigm for smell-based interaction. First, AROMAGEN enables **remote aroma transmission**: users could share olfactory experiences with friends or family in another city by sending a natural language description or image, which the recipient's device reconstructs in real time. Second, AROMAGEN supports **aroma archiving and recall**: users can log descriptions of meaningful aromas—a favorite meal, a childhood kitchen, a travel memory—and relive them on demand, extending the system's role from food simulator to personal olfactory diary. Third, AROMAGEN serves as a **creative tool for interactive experiences**: designers, artists, and storytellers could use language-driven aroma generation to enrich immersive environments, games, or narrative installations without requiring expertise in perfumery or chemistry.

6.3. Limitations and Future Work

The current AROMAGEN prototype is subject to two hardware constraints. First, the wearable dispenser is limited to 12 base odorants, bounding coverage for stimuli outside the

palette. Second, the hardware supports only *sequential* odorant release, precluding true simultaneous blending and limiting compositional richness. Future work will address both dimensions: on the hardware side, we aim to develop dispensing mechanisms capable of simultaneous multi-channel release; on the application side, we plan to extend evaluation beyond food to the broader scenarios described above, assessing the generalizability of language-driven aroma generation across diverse real-world contexts.

7. Conclusion

We presented AROMAGEN, an AI-powered wearable interface that leverages latent olfactory knowledge in multimodal LLMs to generate real-time, general-purpose aromas from free-form text and visual inputs. Through a controlled user study, we demonstrated that AROMAGEN successfully recreates food aromas, with iterative refinement substantially closing the perceptual generation gap. Our results show that zero-shot generation alone matches non-expert human performance, while human-in-the-loop iterative refinement further reduces perceived artificiality to levels comparable to real food. This suggests that language is a viable medium for aroma generation that grows more accurate through human collaboration. We believe that AROMAGEN represents a step towards intelligence-driven olfactory interfaces, opening new possibilities for communication, wellbeing, and immersive technologies.

References

[1] Manuel Aleixandre, Dani Prasetyawan, and Takamichi Nakamoto. Automatic scent creation by cheminformatics method. *Scientific Reports*, 14(1):31284, 2024.

[2] Judith Amores and Pattie Maes. Essence: Olfactory Interfaces for Unconscious Influence of Mood and Cognitive Performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems Pages 28-34*, pages 28–34. ACM Press, 2017.

[3] Judith Amores, Mae Dotan, and Pattie Maes. Development and Study of Ezzence: A Modular Scent Wearable to Improve Wellbeing in Home Sleep Environments. *Frontiers in Psychology*, 13:791768, 2022.

[4] Giada Brianza, Jesse Benjamin, Patricia Cornelio, Emanuela Maggioni, and Marianna Obrist. *QuintEssence:* A Probe Study to Explore the Power of Smell on Emotions, Memories, and Body Image in Daily Life. *ACM Transactions on Computer-Human Interaction*, 29(6):1–33, 2022.

[5] Jas Brooks, Steven Nagels, and Pedro Lopes. Trigeminal-based Temperature Illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Honolulu HI USA, 2020. ACM.

[6] Jas Brooks, Shan-Yuan Teng, Jingxuan Wen, Romain Nith, Jun Nishida, and Pedro Lopes. Stereo-Smell via Electrical Trigeminal Stimulation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Yokohama Japan, 2021. ACM.

[7] Johannes H.F. Bult, Rene A. de Wijk, and Thomas Hummel. Investigations on multimodal sensory integration: Texture, taste, and ortho- and retronasal olfactory stimuli in concert. *Neuroscience Letters*, 411(1): 6–10, 2007.

[8] Eunsol Sol Choi, Yi Xie, Sihao Chen, Elin Carstensdottir, and Edward F. Melcer. Scented Days: Exploring the Capacity of Smell Narratives. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play*, pages 288–293, New York, NY, USA, 2024. Association for Computing Machinery.

[9] David Dobbelstein, Steffen Herrdum, and Enrico Rukzio. inscent: A wearable olfactory display as an amplification for mobile notifications. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 130–137, 2017.

[10] Andrew Dravnieks. Odor quality: Semantically generated multidimensional profiles are stable. *Science*, 218 (4574):799–801, 1982.

[11] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(2): 230–244, 2022.

[12] Dewei Feng, Carol Li, Wei Dai, Alistair Pernigo, Yunge Wen, and Paul Pu Liang. SMELLNET: A Large-scale Dataset for Real-world Smell Recognition, 2025.

[13] Idan Frumin, Ofer Perl, Yaara Endevelt-Shapira, Ami Eisen, Neetai Eshel, Iris Heller, Maya Shemesh, Aharon Ravia, Lee Sela, Anat Arzi, and Noam Sobel. A social chemosignaling function for human handshaking. *eLife*, 4:e05154, 2015.

[14] Richard C. Gerkin. Parsing sage and rosemary in time: The machine learning race to crack olfactory perception. *Chemical Senses*, 46:bjab020, 2021.

[15] Jay A Gottfried. Central mechanisms of odour object perception. *Nature Reviews Neuroscience*, 11(9):628–641, 2010.

[16] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, pages 139–183. Elsevier, 1988.

[17] Sadakichi Hartmann. A Trip to Japan in Sixteen Minutes, 1902.

[18] Keisuke Hasegawa, Liwei Qiu, and Hiroyuki Shinoda. Midair Ultrasound Fragrance Rendering. *IEEE Trans-*

actions on Visualization and Computer Graphics, 24 (4):1477–1485, 2018.

[19] Morton Heilig. Sensorama simulator, 1962.

[20] Rachel S. Herz and Trygg Engen. Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, 3 (3):300–313, 1996.

[21] Richard Hopper, Daniel Popa, Emanuela Maggioni, Devarsh Patel, Marianna Obrist, Basile Nicolas Landis, Julien Wen Hsieh, and Florin Udrea. Multi-channel portable odor delivery device for self-administered and rapid smell testing. *Communications Engineering*, 3 (1):141, 2024.

[22] Thomas Hörberg, Maria Larsson, and Jonas K. Olofsson. The semantic organization of the English odor vocabulary. *Cognitive Science*, 46(11):e13205, 2022.

[23] Hardeep Kaur, C Kishor Kumar Reddy, D Manoj Kumar Reddy, and Marlia Mohad Hanafiah. Single modality to multi-modality: The evolutionary trajectory of artificial intelligence in integrating diverse data streams for enhanced cognitive capabilities. In *Multimodal Generative AI*, pages 297–322. Springer, 2025.

[24] Andreas Keller and Leslie B. Vosshall. Olfactory perception of chemically diverse molecules. *BMC Neuroscience*, 17(1):55, 2016.

[25] Andreas Keller, Richard C. Vosshall, Leslie B. Vosshall, et al. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355 (6327):820–826, 2017.

[26] Murathan Kurfalı, Pawel Herman, Stephen Pierzchajlo, Jonas Olofsson, and Thomas Hörberg. Representations of smells: The next frontier for language models? *Cognition*, 2025.

[27] Brian K. Lee, Emily J. Mayhew, Benjamin Sanchez-Lengeling, Jennifer N. Wei, Wesley W. Qian, Kelsie A. Little, Matthew Andres, Britney B. Nguyen, Theresa Moloy, Jacob Yasonik, Jane K. Parker, Richard C. Gerkin, Joel D. Mainland, and Alexander B. Wiltschko. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023.

[28] Junxian Li, Yanan Wang, Zhitong Cui, Jas Brooks, Yifan Yan, Zhengyu Lou, and Yucheng Li. Mid-Air Gestures for Proactive Olfactory Interactions in Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18, Yokohama Japan, 2025. ACM.

[29] Yucheng Li, Yanan Wang, Mengyuan Xiong, Max Chen, Yifan Yan, Junxian Li, Qi Wang, and Preben Hansen. AromaBite: Augmenting Flavor Experiences Through Edible Retronasal Scent Release. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, Yokohama Japan, 2025. ACM.

[30] Paul Pu Liang. Foundations of multisensory artificial intelligence. *arXiv preprint arXiv:2404.18976*, 2024.

[31] Paul Pu Liang, Karan Ahuja, and Yiyue Luo. Multimodal ai for human sensing and interaction. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2025.

[32] Yang Liu, Weixing Chen, Yongjie Bai, Xiaodan Liang, Guanbin Li, Wen Gao, and Liang Lin. Aligning cyber space with physical world: A comprehensive survey on embodied ai. *IEEE/ASME Transactions on Mechatronics*, 2025.

[33] Asifa Majid. Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences*, 25(2):111–123, 2021.

[34] Asifa Majid and Niclas Burenhult. Odors are expressible in language, as long as you speak the right language. *Cognition*, 130(2):266–270, 2014.

[35] Daiki Mayumi, Yugo Nakamura, Yuki Matsuda, and Keiichi Yasumoto. BubblEat: Designing a Bubble-Based Olfactory Delivery for Retronasal Smell in Every Spoonful. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, Yokohama Japan, 2025. ACM.

[36] Siddharth Mehrotra, Anke Brocker, Marianna Obrist, and Jan Borchers. The Scent of Collaboration: Exploring the Effect of Smell on Social Interactions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, New Orleans LA USA, 2022. ACM.

[37] Severino Muñoz-Aguirre, Akihito Yoshino, Takamichi Nakamoto, and Toyosaka Moriizumi. Odor approximation of fruit flavors using a QCM odor sensing system. *Sensors and Actuators B: Chemical*, 123(2): 1101–1106, 2007.

[38] Takamichi Nakamoto. Olfactory Display and Odor Recorder. In *Essentials of Machine Olfaction and Taste*, pages 247–314. Wiley, 1 edition, 2016.

[39] T. Nakamoto, M. Ohno, and Y. Nihei. Odor Approximation Using Mass Spectrometry. *IEEE Sensors Journal*, 12(11):3225–3231, 2012.

[40] Bart M. Nicolaï, Evelien Micholt, Nico Scheerlinck, Thomas Vandendriessche, Maarten L.A.T.M. Hertog, Ian Ferguson, and Jeroen Lammertyn. Real time aroma reconstruction using odour primaries. *Sensors and Actuators B: Chemical*, 227:561–572, 2016.

[41] Dani Prasetyawan and Takamichi Nakamoto. Odor Reproduction Technology Using a Small Set of Odor Components. *IEEJ Transactions on Electrical and Electronic Engineering*, 19(1):4–14, 2024.

[42] Nimesha Ranasinghe, Chow Eason Wai Tung, Ching Chiuan Yen, Ellen Yi-Luen Do, Pravar Jain, Nguyen Thi Ngoc Tram, Koon Chuan Raymond Koh,

David Tolley, Shienny Karwita, Lin Lien-Ya, Yan Liangkun, and Kala Shamaiah. Season Traveller: Multisensory Narration for Enhancing the Virtual Reality Experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM Press, 2018.

[43] Nimesha Ranasinghe, Koon Chuan Raymond Koh, Nguyen Thi Ngoc Tram, Yan Liangkun, Kala Shamaiah, Siew Geuk Choo, David Tolley, Shienny Karwita, Barry Chew, Daniel Chua, and Ellen Yi-Luen Do. Tainted: An olfaction-enhanced game narrative for smelling virtual ghosts. *International Journal of Human-Computer Studies*, 125:7–18, 2019.

[44] Aharon Ravia, Kobi Snitz, Danielle Honigstein, Maya Finkel, Rotem Zirler, Ofer Perl, Lavi Secundo, Christophe Laudamiel, David Harel, and Noam Sobel. A measure of smell enables the creation of olfactory metamers. *Nature*, 588(7836):118–123, 2020.

[45] Lisa Reichl and Martin Kocur. Investigating the Impact of Odors and Visual Congruence on Motion Sickness in Virtual Reality. In *30th ACM Symposium on Virtual Reality Software and Technology*, pages 1–12, Trier Germany, 2024. ACM.

[46] Celine E Riera, Eva Tsaousidou, Jonathan Halloran, Patricia Follett, Oliver Hahn, Mafalda MA Pereira, Linda Engström Ruud, Jens Alber, Kevin Tharp, Courtney M Anderson, et al. The sense of smell impacts metabolic health and obesity. *Cell metabolism*, 26(1): 198–211, 2017.

[47] Clemens Schartmüller and Andreas Riener. Sick of Scents: Investigating Non-invasive Olfactory Motion Sickness Mitigation in Automated Driving. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 30–39, Virtual Event DC USA, 2020. ACM.

[48] Kobi Snitz, Adi Yablonka, Tali Weiss, Idan Frumin, Rehan M. Khan, and Noam Sobel. Predicting Odor Perceptual Similarity from Odor Structure. *PLoS Computational Biology*, 9(9):e1003184, 2013.

[49] David R Thomas. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2):237–246, 2006.

[50] Yanan Wang, Judith Amores, and Pattie Maes. On-Face Olfactory Interfaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–9, Honolulu HI USA, 2020. ACM.

[51] Takao Yamanaka, Ryosuke Matsumoto, and Takamichi Nakamoto. Study of recording apple flavor using odor recorder with five components. *Sensors and Actuators B: Chemical*, 89(1-2):112–119, 2003.

[52] T. Yamanaka, B. Wyszynski, and T. Nakamoto. Study of odor recorder for recording recipe of orange flavor. In *TRANSDUCERS '03. 12th International Conference on Solid-State Sensors, Actuators and Microsystems. Digest of Technical Papers (Cat. No.03TH8664)*, pages 1140–1143, Boston, MA, USA, 2003. IEEE.

[53] Yasuyuki Yanagida, Haruo Noma, Nobuji Tetsutani, and Akira Tomono. An unencumbering, localized olfactory display. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems - CHI '03*, page 988, Ft. Lauderdale, Florida, USA, 2003. ACM Press.

[54] Leijing Zhou, Yiqing Zhang, Xin An, and Junxian Li. E-scent Coach: A Wearable Olfactory System to Guide Deep Breathing Synchronized with Yoga Postures. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–14, Bordeaux/Talence France, 2025. ACM.

A. Expert Interview Protocol

Expert Background

- How many years of studies or practice do you have in this field?
- What is your specialization (e.g., perfumer, flavor chemist, olfactory artist)?

Temporal and Perceptual Justification

- What are the perceptual differences between simultaneous mixture and sequential release? Can sequential release approximate the perception of a mixture?
- Is the use of essential oils a common practice in fragrance blending applications?
- What is an appropriate duration for users to perceive and reflect on a smell?
- How many active ingredients are needed to reach a perceptual "sweet spot" complex enough to be realistic, but not overwhelming?

Aroma Space and Palette Selection

- How many base odorants are needed to achieve meaningful coverage of everyday food aromas?
- Given the perceptual categories identified in our formative study (sweet, savory, sour, burnt/smoked, fresh), does our selection of base odorants provide adequate coverage?
- What is the role of volatility in aroma perception? Please confirm or revise the volatility scores and characteristic notes for our base odorants.

B. System Prompts

Zero-Shot Generation Prompt

```
You are AromaAI, a food smell recreator. Given a food
     or beverage description, output a ratio vector
     over 12 base odorants that recreates the smell as
     recognizably as possible.

Response must be a single valid JSON object. No
     markdown, no extra text.
```

```
ODORANT PALETTE:
{{ scents_json }}

Attributes:
- volatility: higher = brighter, faster-presenting top
     note
- note: qualitative character (e.g. "cheesy, sour",
     "fruity, sweet", "spicy hotpot")
- location: hardware index only, no semantic meaning

TASK:
1. Identify the food's dominant smell identity from the
     'note' field.
2. Extract food beats (food category, preparation
     state, key notes) - do NOT invent ingredients the
     user did not mention.
3. Allocate ratios: prefer high-volatility odorants for
     primary recognition; use 3-6 active odorants, set
     the rest to 0.00.
4. In justification, list the food beats and explain
     ratio choices per beat.

CONSTRAINTS (strict):
- All 12 odorants must appear in output, even if 0.00
- Ratios sum to exactly 1.00, each >= 0, two decimal
     places
- Use smell names exactly as in the palette

OUTPUT:
{
  "scent_ratios": {
    "<scent_name>": <ratio>,
    ...
  },
  "justification": "<food beats + ratio reasoning>"
}
```

Iterative Refinement Prompt

```
You are AromaAI in REVISION mode. Refine the current
     ratio vector based on user feedback to better
     match the target food smell.

Response must be a single valid JSON object. No
     markdown, no extra text.

ODORANT PALETTE:
{{ scents_json }}

Attributes:
- volatility: higher = brighter top note
- note: qualitative character cues
- location: hardware index only

REVISION RULES (in priority order):
1. Address the latest feedback - interpret it as what
     is missing, too strong, or wrong.
2. Anchor unchanged odorants - preserve ratios the user
     did not criticize.
3. Targeted changes only - adjust only what the
     feedback demands.
4. Rebalance so ratios sum to exactly 1.00.
5. Prefer shifting existing ratios over introducing new
     odorants.

CONSTRAINTS (strict):
- All 12 odorants must appear in output, even if 0.00
- Ratios sum to exactly 1.00, each >= 0, two decimal
     places
- Use smell names exactly as in the palette

USER MESSAGE FORMAT:
ORIGINAL REQUEST: initial food target
CURRENT RATIOS: ratio vector to revise
PRIOR FEEDBACK HISTORY: previous rounds (may be empty)
>>> LATEST FEEDBACK <<<: your primary instruction

OUTPUT:
{
  "scent_ratios": {
    "<scent_name>": <ratio>,
    ...
  },
  "justification": "<how this revision improves food
     similarity; what feedback was addressed>",
  "changes_made": "<which ratios increased / decreased
     / zeroed / introduced and why>"
}
```

C. Odorant Sources

See Table 5.

D. Odorant Material Types

Essential Oil Natural volatile oil directly extracted from plant material via steam distillation or cold pressing, retaining the plant's characteristic aroma with complex chemical composition.

Tincture Plant extract prepared by soaking botanical material in alcohol, yielding a diluted aromatic liquid with a milder aroma profile than essential oils.

Fragrance Oil Synthetically compounded aromatic blend formulated from multiple aroma chemicals, offering consistent and stable aroma performance independent of natural plant sources.

Pure Aroma Chemical Single isolated chemical compound of defined molecular identity, used as a raw material in fragrance formulation.

Powder Dry, finely ground plant material that releases aroma passively, with lower volatility and less consistent aroma diffusion compared to liquid extracts.

E. Hardware Specification

Neckwear Digital Scent Player (NECK BAND)		
Hardware	Brand	SCENTAC
	Model	X-Scent 3.0/ Olfex One
	Product Name	Digital Scent Player
	Function	Synchronized scent player with movie (Video/Game)scenes, providing viewers/player with a rich olfactory experience to enhance the overall movie-watching/game playing experience.
	Color	Black
Basic Info	Device dimension	195mm*245mm*56mm
	Weight	＜380g
	Device Material	ABS、Silicone
	Max Scents	12
Scent Cartridge	Size	92*22*35mm
	Qty	2
	Scent Max	12
	Material	ABS
Interface	Power interface	Magnetic suction interface
	Data interface	433M/ Bluetooth
Power Info	Battery Capacity	1000mAh/3200mAh
	Battery removable	No
	Input	DC5V
	Max power	2W

Figure 14. Hardware Specification Sheet

Odorant	**Type**	**Brand**	**Ingredients / Formula**
Cumin	Essential Oil	Silky Scents	100% pure Cumin seed oil (*Cuminum cyminum*)
Eucalyptus	Essential Oil	Creative Flavours & Fragrances	Eucalyptus essential oil (*Eucalyptus* spp.)
Onion	Powder	Badia	Onion powder
Ylang Ylang	Essential Oil	White Naturals	Pure Ylang Ylang essential oil; 1 fl oz
Sichuan Oil (Linalool)	Essential Oil	Creative Flavours & Fragrances	High concentration of Linalool derived from Sichuan peppercorns, providing a woody, floral, and citrus-spicy top note.
Red Clover	Tincture	Cambridge Naturals	Organic Red Clover flowering tops (*Trifolium pratense*); grain alcohol (45–55% by volume); deionized water; herb strength ratio 1:3
Cypress	Essential Oil	Perfumer's Apprentice	*Cupressus Sempervirens* (Spain). Smoky, sweet-balsamic odor with fresh pine, woody, earthy, dry spicy cedar character. CAS#: 8013-86-3
Thyme	Tincture	Cambridge Naturals	Organic Thyme leaf (*Thymus vulgaris*); grain alcohol (45–55% by volume); deionized water; herb strength ratio 1:2
Strawberry	Fragrance Oil	P&J	Fragrance oil blend (essential oils, aroma chemicals, carrier oils; proprietary composition)
Isovaleric Acid	Pure Chemical	Perfumer's Apprentice	Cheese, dairy, acidic, sour, pungent, fruity, fatty. Use level: $\leq$1.00% solution. CAS#: 503-74-2
Cinnamon	Essential Oil	Frontier Co-op	Organic sunflower oil; organic cinnamon oil
Sage	Essential Oil	Cambridge Naturals	100% pure Sage essential oil

Table 5. Sources and compositions of the 12 base odorants used in AROMAGEN's palette.