Towards Understanding the Design of Mixed Reality Systems to Enrich the Beverage Experience

Yuchen Zheng
Exertion Games Lab, Department of Human-Centred Computing
Monash University
Melbourne, Australia
yuchen.zheng@monash.edu

Zhuo Wang
Department of Computing
Xi'an Jiaotong-liverpool University
Su Zhuo, China
zhuohci01@gmail.com

Hongyue Wang
Exertion Games Lab, Department of Human-Centred Computing
Monash University
Melbourne, Australia
hongyue@exertiongameslab.org

Don Samitha Elvitigala
Exertion Games Lab, Department of Human-Centred Computing
Monash University
Melbourne, Australia
don.elvitigala@monash.edu

Florian 'Floyd' Mueller
Exertion Games Lab, Department of Human-Centred Computing
Monash University
Melbourne, Australia
floyd@exertiongameslab.org

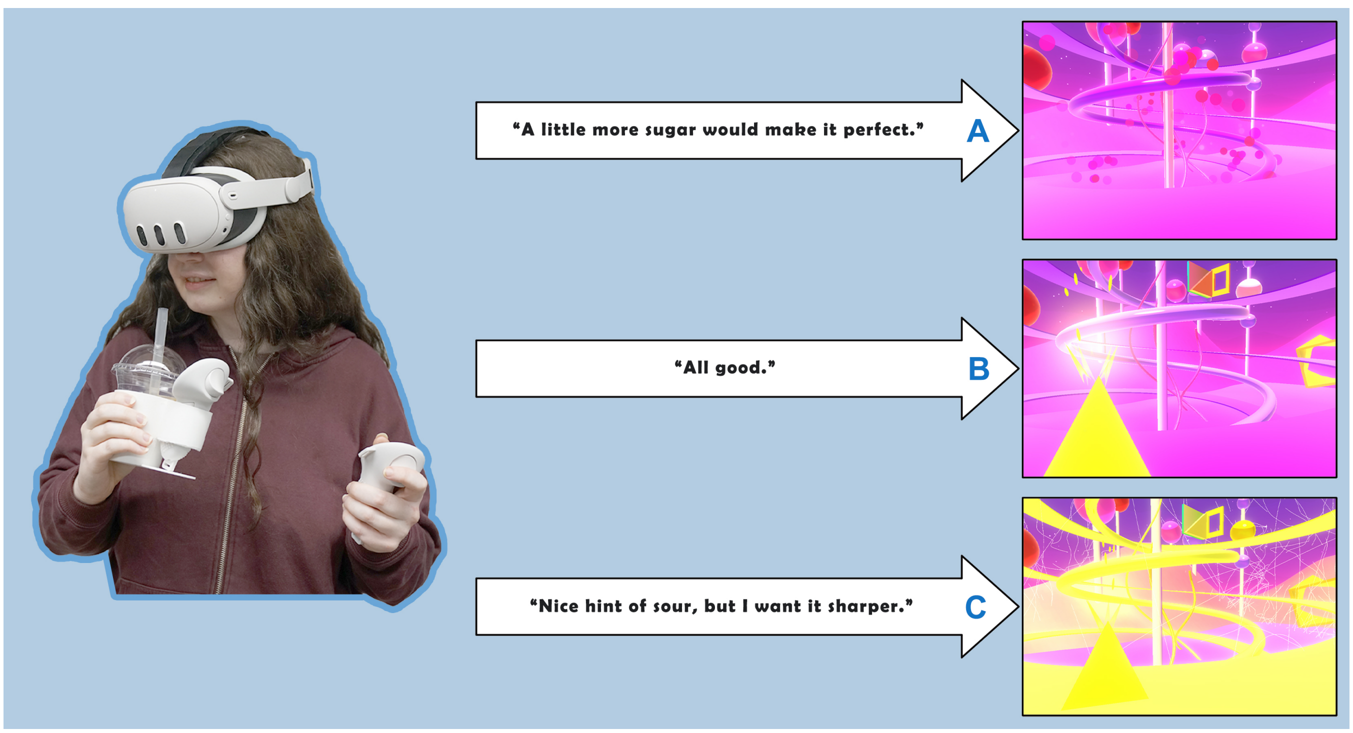


Figure 1: Overview of XTea system. Real-time modulation of the virtual environment based on participant input during bubble tea consumption: A) request for increased sweetness; B) satisfaction with current state; C) request for heightened sourness.

Abstract

Drinking is an inherently multisensory activity, yet the potential of immersive technology to dynamically shape flavor experiences remains underexplored in Human-Food Interaction (HFI) research. We introduce "XTea", an adaptive beverage cup-based system that integrates large language models to translate natural language input into modifications of a parameterized immersive environment experienced through a headset when drinking bubble tea. Through

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790615

a study with 12 bubble tea enthusiasts, we derived themes that demonstrate how "XTea" can enrich sensory engagement, support personalized and agentic experiences, and foster social qualities of drinking, pointing toward new explorations for multisensory HFI design. We also present four design strategies for multisensory beverage experiences. Ultimately, we aim to contribute to the advancement of HFI research on how multisensory interaction design can enrich flavor perception and engagement.

CCS Concepts

• **Human-centered computing → Interaction design**.

Keywords

Multisensory, Human-Food Interaction, Mixed Reality, Virtual Reality, Flavor Perception

ACM Reference Format:
Yuchen Zheng, Zhuo Wang, Hongyue Wang, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2026. Towards Understanding the Design of Mixed Reality Systems to Enrich the Beverage Experience. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3772318.3790615

1 INTRODUCTION

Drinking is a multisensory experience shaped by the interplay of taste, smell, vision, and sound [79, 81]. People drink beverages not only to quench thirst, but also to enjoy sensory pleasure, take emotional breaks, and connect with others [27, 43, 87]. Prior studies in Human-Food Interaction (HFI) [35, 44] have explored how interactive technology can help craft multisensory drinking experiences, from shape-changing drinks through cymatics tableware [17], to gustosonic drinking experiences through interactive straws [87], yet considerable potential within HFI for exploring richer multisensory drinking experiences remains unexplored [53, 87]. In particular, we find that the potential of immersive technology to dynamically shape the act of drinking by embedding flavor-congruent visual and auditory cues in the immersive environment is underexplored.

Charles Spence, who has extensively studied the multisensory foundations of flavor perception, said: *"To date, very few researchers have taken seriously the suggestion that a richer spatial dimension is present in flavour experiences, as compared to in the majority of other perceptual experiences"*[67], highlighting the importance of spatiality in shaping flavor experiences through (interaction) design. Although research has shown that flavor perception is highly context-dependent [64], and influenced by what we see [54, 67, 69, 72], hear [20, 85], and do [16, 46], little is known about how multisensory cues in immersive virtual environments can interact with embodied actions to enrich the experience. At the same time, the proliferation of natural-language–based intent recognition has introduced new opportunities for interpreting user goals and contextual cues in interactive food-related applications. Prior work includes robots that infer cooking intentions from natural-language recipes [8] and speech-based food journals that identify in-situ eating intentions from free-form speech [37]. However, within HFI, little attention has been paid to how natural language input can shape system's multisensory responses. One mechanism for implementing of natural-language–based intent recognition—large language models (LLMs)—has rapidly advanced in recent years, opening new possibilities for real-time personalization in interactive food-related applications, such as personalized food recommendation [89] and nutrition-aware systems [91]. However, little attention has been paid to how natural-language-based intent recognition might help to enrich personalized flavor experiences.

We argue that natural-language input, which is casual, intuitive, and expressive, holds potential for supporting adaptive multisensory environments that respond to individual preferences in real time. Here, we build on the core idea of adaptive user interfaces [11, 34] in which a system changes itself to meet user' needs, and applies it to a responsive environment [13]. We define "adaptive" as the ability to continuously modulate sensory parameters of the environment based on users' input. In immersive contexts, such adaptivity could allow users to co-create their own drinking experience, enriching consumption into a playful, engaging, and personalized experience. To investigate this opportunity, we begin answering the following research question: "How do we design enriched drinking experiences using immersive technology that leverages embodied actions and adaptive visual and auditory cues?"

We introduce "XTea", a multisensory Mixed Reality (MR) [62] system that bridges the physical act of drinking real bubble tea with an adaptive virtual environment, creating an Augmented Virtuality (AV) [39] experience. We chose bubble tea because it is a trendy beverage with a rich and playful character [33], which makes it well suited for multisensory research [92]. The system consists of an LLM-powered adaptive virtual environment (the virtual component) and a custom-designed "XTea" cup (the physical component), integrating the physical act of drinking in the real world with corresponding virtual feedback into the virtual environment (the physically augmented virtual component). This system enables participants to shape the virtual environment in real time by interpreting natural language input to adjust visual and auditory parameters (Fig. 1A-1C).

Overall, this work makes the following contributions to the fields of HFI:

(1) The XTea system: We introduce "XTea", a multisensory beverage drinking system that integrates visual, auditory, and physical components. We believe this system can provide inspiration for HFI researchers to design future multisensory flavor experiences.
(2) Thematic findings: Our findings highlight the types of sensory elements that are most effective and show how the adaptive system can support personalized flavor modulation. We hope these findings inform researchers interested in designing immersive drinking systems.
(3) Design strategies: We propose four design strategies for future multisensory beverage experiences. These strategies are aimed to be actionable guidance for practitioners and designers seeking to create playful and engaging food experiences.

2 RELATED WORK

We now summarize previous work that inspired our research.

2.1 The use of immersive technology in multisensory HFI

2.1.1 Multisensory design. Taste perception is inherently multisensory, arising from the integration of taste in the brain with visual, olfactory, tactile, and auditory cues [64, 68]. Recognizing this, researchers in HFI have explored technologies that leverage crossmodal interactions to enhance eating and drinking experiences [35, 42]. A key area of focus within this domain is Multisensory Human-Food Interaction (MHFI) [81], which examines how sensory stimuli shape food perceptions and overall dining experiences [19, 67]. A notable example is "MetaCookie+" [45], an augmented reality system that overlays visual and olfactory cues onto a plain cookie, inducing illusory flavors without altering its chemical composition. By synchronizing visuals and scents (e.g., chocolate), users perceive the plain cookie as flavored, demonstrating how incongruent but strategically aligned sensory inputs can override actual taste. Similarly, the "Vocktail" [54] system uses an LED to change the drink's color, micro-pumps to release scented vapors, and electrodes on the rim to stimulate taste buds. By aligning color, scent, and taste cues, it transforms water into convincingly flavored drinks. These works are important because they demonstrate how crossmodal cues can fundamentally reshape flavor perception, establishing MHFI as a domain that reveals both the theoretical mechanisms of multisensory integration and practical opportunities for designing novel food experiences. However, these approaches lack dynamic or immersive interactions.

The advent of Virtual Reality (VR) technology, has expanded the horizon of MHFI, providing researchers with vision-based headsets to design dynamic and immersive environments that can influence sensory responses [66, 78]. Such headset-based systems can create a sense of presence that emotionally connects users to virtual settings, enabling them to experience food and beverages in ways that transcend physical and sensory constraints [77]. For instance, in the study by Wang et al., VR technology was used to manipulate the appearance of coffee, such as adding "virtual milk" by changing its color, which influenced participants' flavor perception without altering the physical beverage itself [86]. Similarly, Ammann et al. [1] employed VR technology to present an emotionally aversive scene—depicting a dog defecating chocolate, which significantly reduced participants' willingness to consume real chocolate, illustrating the capacity of VR technology to shape eating behavior through sensory incongruence. Separately, studies on social drinking in virtual reality have highlighted how immersive environments can alter perceived intoxication and drinking behavior, calling for a more safety-conscious, context-aware VR design [15]. These works show that VR technology can play a role in MHFI by offering immersive and emotionally engaging environments that reshape flavor perception and eating behavior beyond physical constraints. However, most VR systems in MHFI remain non-adaptive and designer-driven, meaning that participants have little opportunity to actively shape their own experiences.

Across existing MHFI studies, current work remains predominantly designer-driven and misses a consideration of the participant involvement as a meaningful source of input [1, 45, 54, 86]. These works treat the participant's interaction merely as a trigger [41]. Even in the most participant-driven work among them, "Vocktail" [54], the user's bodily involvement is tongue contact that simply triggers electrodes rather than operating as an expressive medium; additionally, participant input is configured through a mobile application rather than emerging from the participant's own embodied actions [23] as an active, meaning-producing component of the experience. This reduction in participant involvement may limit the movement of MHFI toward a more humanized technological future [41]. We propose that verbal input, an essential modality of embodied communication in human–computer interaction [58], offers a pathway for MHFI work to move beyond designer-driven structures toward participant-driven meaning-making, fostering engagement grounded in the participant's own expressive input. This proposal may help articulate, at a conceptual level, how verbal input can participate in the emergence of engagement within MHFI.

2.1.2 Bubble tea. We believe that bubble tea offers a particularly rich context for applying multisensory design because of its diverse sensory attributes, such as the chewiness of pearls, the layering of flavors, and the interplay between hot and cold elements [33, 92]. Previous research on bubble tea and multisensory design has focused mainly on passive aspects, such as packaging design that combines visual, tactile, and aromatic elements to enhance consumer enjoyment and emotional satisfaction before consumption [65, 80]. However, there has been little exploration of how interactive multisensory design could transform the drinking experience itself [92]. We believe that with its layered textures and dynamic flavor profiles, bubble tea provides an ideal opportunity to investigate this potential.

2.1.3 Crossmodel correspondences. Crossmodal correspondences refer to the associations people make between features in different sensory modalities [22], such as associating high-pitched sounds with sour tastes or the color pink with sweetness [69].

An increasing number of studies use VR technology in crossmodal research [6, 7, 50, 61], recognizing its immersive qualities such as presence, engagement, and realism. These features make VR technology a cost-effective alternative to real-world settings. For instance, Chen et al. [18] created three virtual environments to examine whether congruent and incongruent external visual cues presented in virtual environment would influence the perception of sweetness and product liking. Likewise, Bangcuyo et al. [3] developed an immersive virtual café filled with visual, auditory, and olfactory cues typical of such settings, in order to investigate differences between hedonic data collected in the virtual café and those obtained elsewhere, thereby evaluating whether virtual environments could serve as more reliable predictors of future coffee preferences. Most of these investigations have focused on psychology, consumer science, and marketing, while comparatively little attention has been paid to applying such theoretical findings to the design of interactive consumption systems, particularly those aimed at enriching playful user experiences.

2.2 The potential of responsive environments in shaping dining experiences

Adaptiveness has long been explored at the intersection of architecture and human–computer interaction (HCI), where environments respond to inhabitants by dynamically reshaping material or spatial configurations—through lighting, kinetic surfaces, or airflow—in response to human behavior or changing conditions [13]. For example, Aegis Hyposurface [1] features a kinetic skin that reacts to movement and sound, transforming the façade into a performative and participatory interface. Similarly, Toyo Ito's Tower of Winds [2] responds to city sounds, wind, and light through a rhythmic play of lights. These examples show that adaptiveness is not only functional—it invites interactivity and play. Such systems reframe architecture as more than passive enclosure; they become perceptual and experiential centers. This view aligns with architectural phenomenology [48], where space is continuously sensed, shaped, and reinterpreted through interaction.

The idea that environments can dynamically shape human perception also resonates with findings in sensory science that have shown that environmental elements such as color [31], lighting [12], atmosphere [88], and acoustics [85] can influence how we perceive meals. For example, Bschaden et al. [12] conducted two controlled experiments in which they manipulated the intensity of the lighting and the presence of table linen, and demonstrated that lower lighting levels combined with tablecloths can increase food intake, prolong meal duration, and enhance sensory evaluations of the food. These effects are associated with crossmodal theories, as previously discussed, and suggest that dining is not solely a gustatory process, but is shaped by our surrounding environment. We draw inspiration from these insights and ask: If static environmental features can influence taste perception, what new forms of sensory engagement might be possible in adaptive environments—especially those capable of responding dynamically to the diner? We believe that adaptive immersive environment offers a promising design space for creating multisensory experiences that are not only immersive, but also capable of responding to users' immediate preferences. There is a potential to reframe the virtual environment as a kind of liquid architecture [60] in which sensory parameters such as color, sound, and motion adaptively reshape the user's beverage experience.

This potential motivates our investigation into how adaptive immersive environments can be used to enrich the bubble tea experience.

2.3 Growing use of LLMs for adaptive immersive environment

LLMs have been rapidly growing in both capability and application and have increasingly been used in adaptive VR headset-based environments to enable dynamic interactions [2, 10], extending from entertainment [4, 55] to social applications [14, 59, 82] and health treatments [71]. Such systems provide real-time problem-solving and adapt continuously to user input. A vivid example comes from Dungeons & Dragons, in which an artificial intelligence (AI) Dungeon Master powered by ChatGPT acts as a virtual facilitator, generating coherent narratives and dynamically responding to player input, potentially enhancing narrative adaptability and gameplay freedom [93]. Here, "adaptive" is typically defined as the capacity of such environments to continuously adjust their output in response to user input [2, 10, 14, 82, 93]. We argue that, in immersive settings, this "output" can be extended to the multisensory properties of the environment, which also show possibilities for situating adaptive immersive systems within the domain of HFI (as we discussed in last section). Meanwhile, LLM-driven applications in HFI have primarily focused on sustainable food systems [73], personalized food recommendations [40, 89], and nutrition-aware systems [90, 91]. These advancements, however, often remain constrained to health-related domains, with relatively little attention given to supporting the user experience, such as fostering a playful and engaging beverage experience, which is what we complement with our work.

2.4 Summary

While existing work in MHFI and crossmodal perception has revealed how sensory cues shape flavor experiences, most systems remain static and designer-driven, lacking adaptivenessand opportunities for embodied actions that support meaningful participant involvement. Research on adaptive systems highlights the potential to dynamically shape human perception but has rarely been applied to dining contexts. Similarly, LLMs have been used in functional domains like health, but not for playful beverage experiences. This is probably due to a limited understanding of how adaptive immersive environments might be designed to enrich multisensory flavor perception and drinking experiences. We explore this opportunity by designing and studying an adaptive system that aims to enrich the multisensory beverage experience. Hence, we ask the research question: "How do we design enriched drinking experiences using immersive technology that leverages embodied actions and adaptive visual and auditory cues?"

3 DESIGNING XTEA

"XTea" is an interactive immersive multisensorial system consisting of two key components: an adaptive virtual environment delivered via Meta Quest 3 headset (Fig. 2A) and the "XTea" cup (Fig. 2B). Participants wear the headset to enter the virtual environment, using the left-hand controller (Fig. 2C) for navigation (via thumbstick) and voice input (via "X" and "Y" buttons). In their right hand, they hold the "XTea" cup, a 3D-printed cup that integrates the right-hand controller with a cup of bubble tea to allow tracking of its physical position in the real world. As participants lift the "XTea" cup to drink, this action is mirrored in the virtual environment (Fig. 2D).

We call our adaptive virtual environment "Flavor Drift", as it is capable of dynamically modulating multisensory cues in response to users' verbal input by changing the virtual environment. The name aims to emphasize a continuous, fluid transformation of sensory features in real time.

While the system is implemented using VR technology and could reasonably be described as a VR system, we adopt the definition from the "Reality–Virtuality continuum" [39, 62] and therefore refer to it as a MR system. This framing highlights that the experience of "XTea" emerges from the integration of virtual visual–auditory stimuli with real physical, gustatory, and embodied interactions. We use the term "MR" to foreground discussions in HFI that extend

[1] https://parametrichouse.com/hyposurface/
[2] https://www.architecturelab.net/tower-of-winds-toyo-ito-and-associates/

beyond display technology to include the interplay between digital systems and material, sensory, and bodily dimensions of the drinking experience.

3.1 Interaction

The "XTea" cup (Fig. 2B) was designed to resemble a standard bubble tea cup, supporting common lifting and drinking motions. Its virtual counterpart mirrored these movements in real time (Fig. 2D), maintaining matching dimensions and structure to help participants locate the straw effortlessly. This synchronization was intended to enhance the congruency between physical and virtual experiences.

While holding the "XTea" cup in their right hand, participants could press and hold the "Y" button on left hand controller to voice their input on flavor perception (e.g., "make it less sour"), and press "X" to send the command. Their input was transcribed by Meta Voice SDK and displayed as floating text in front of them within the virtual environment for immediate confirmation. If the transcription did not match their intention, they could press Y again to re-record. To encourage regular reflection and engagement, the system prompts participants every two minutes with a voice message (e.g., "Please take a sip of your drink now and enjoy the moment" or "How does your drink taste right now?"). These prompts served as gentle nudges to elicit verbal responses about the user's current sensory experience or their desired adjustments.

3.2 Virtual environment

The virtual environment, called "Flavor Drift", was crafted with visual and auditory cues and built using Unity, with interactions in virtual environment managed through the XR Interaction Toolkit. The auditory stimuli, which consisted of two soundtracks specifically designed to evoke the sensations of sweetness and sourness [3](used with the composer's permission). We chose these soundtracks as prior work [85] demonstrated in previous crossmodal studies that they can effectively evoke intended taste associations.

We selected two contrasting yet flavor-relevant dimensions for bubble tea—sweetness and sourness—and extracted their associated design features, embedding them as visual elements in our virtual environment. Elements beginning with "SW" represent Sweet-associated components (Fig. 3), while those beginning with "SO" represent Sour-associated components (Fig. 4).

On top of this flavor-based labeling, we layered an additional categorization based on perceptual salience. A human's perceptual system filters information by distinguishing between "ambient" elements that subtly influence perception and "focal" elements that have a direct impact when attention is paid to them [74]. We believe this ambient–focal layered perception also applies here. Therefore, based on their influence on perception, we categorized our "SW" elements into three types: SWA (Sweet Ambient), SWAF (Sweet Ambient–Focal), and SWF (Sweet Focal):

- **SWA:** Elements with ambient influence on sweet perception
- **SWAF:** Elements contributing to both ambient and focal influence on sweet perception
- **SWF:** Elements with focal influence on sweet perception

[3]https://jialindeng.wixsite.com/whispery-savoury/

We also categorized "SO" elements into three types: SOA (Sour Ambient), SOAF (Sour Ambient-Focal), and SOF (Sour Focal):

- **SOA:** Elements with ambient influence on sour perception
- **SOAF:** Elements contributing to both ambient and focal influence on sour perception
- **SOF:** Elements with focal influence on sour perception

Our "SW" and "SO" elements were designed on the basis of established crossmodal studies, which indicate that sweet tastes are visually linked to rounded or curved forms, slow movements, and pink-to-red hues [18, 32, 67], while sour tastes are associated with sharp and angular shapes, fast movements, and yellow-green tones [18, 47, 67]. Building on these findings, we designed all visual elements with the aim of enriching the corresponding perception. These elements form the foundational design of the "Flavor Drift" setting, though additional elements may be incorporated in the future. Next, we articulate how we translated the crossmodal findings into design decisions for the "SW" and "SO" elements in "Flavor Drift" across color, shape, and motion, as well as their corresponding audio components.

3.2.1 "SW" visual elements. The sweet-associated environment (SW) consists of three ambient (SWA), two ambient–focal (SWAF), and four focal elements (SWF) that together construct the visual qualities linked to sweetness. Ambient elements (SWA) comprise the blooming post-processing (SWA1), pink directional light (SWA2), and smooth curved landscape (SWA3). SWA1–3 work together to build a sweet atmosphere: a soft pink glow created by blooming post-processing, matching directional light, and smooth curved landscapes with pink gradients establishes a sweet visual base for the entire scene, grounded in crossmodal visual–taste research (see Table 2).

Sweet-associated elements that function as both ambient and focal (SWAF) comprise a spiral pathway (SWAF1) and decorative objects along it (SWAF2). They both remain static yet shift from focal to ambient with viewer proximity, embodying crossmodal links between sweetness, rounded forms, and pink hues (detailed in Table 3).

Focal sweet-associated elements (SWF1–SWF4) emphasize sweetness through glossy candy-like spheres, lollipop-like structures, and particle systems arranged along the spiral path. Rounded geometries with reflective pink-red shader, and pink-red particle systems collectively consistent with crossmodal visual–taste research on sweetness (detailed in Table 4).

3.2.2 "SO" visual elements. The sour-associated environment (SO) consists of one ambient (SOA), one ambient–focal (SOAF), and six focal elements (SOF) that together construct the visual qualities linked to sourness. The ambient component (SOA1) is an intense yellow directional light that floods the entire scene, drawing on crossmodal findings linking yellowness to sour taste experiences (Table 5). The ambient–focal element (SOAF1) consists of high-speed yellow Visual Effects (VFX) stripes whose rapid motion initially attracts attention before gradually blending into the environment, establishing a sharp motion profile associated with sourness and grounded in crossmodal visual–taste research (see Table 6).

Table 1: Components of the XTea system and their roles within Mixed Reality.

Component	Type	Function in the System	MR classification
Flavor Drift	Virtual	Provides the immersive visual–auditory scene whose parameters (color, geometry, motion, sound) can be adapted.	Main virtual space; by itself it would be a VR environment.
"XTea" cup physical component (Fig. 2B)	Physical	Real cup used for holding and drinking bubble tea; provides haptic and gustatory sensations and real drinking actions.	Real-world sensation and bodily action.
"XTea" cup virtual component (Fig. 2D)	Mixed	Links physical drinking movement to virtual feedback: the virtual cup mirrors physical movement.	This coupling of real actions and adaptive virtual responses is what makes XTea an Augmented Virtuality system within Mixed Reality, even though it is implemented with VR hardware.

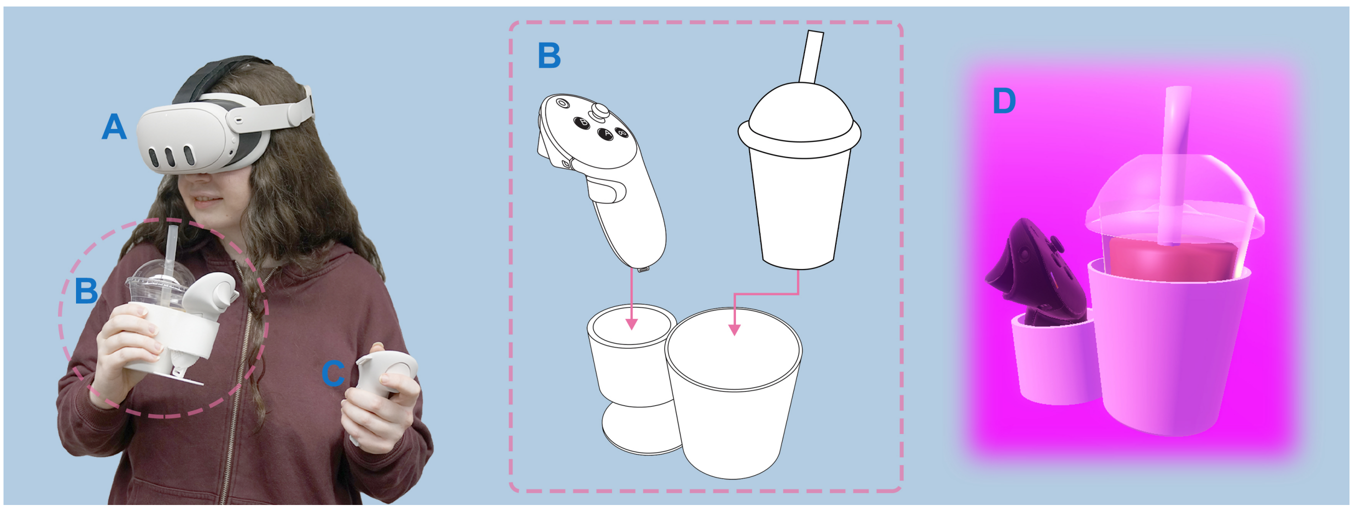


Figure 2: Components of XTea cup: A) VR headset; B) XTea cup; C) left hand controller; D) virtual counterpart of the XTea cup rendered within the virtual environment.

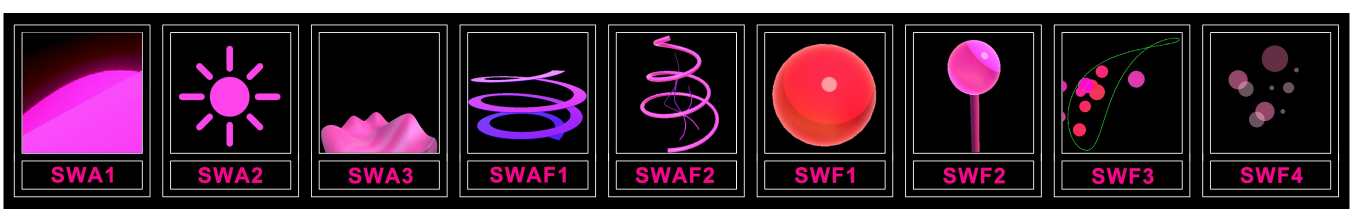


Figure 3: "SW" elements.

Focal sour-associated elements (SOF1–6) extend this language through angular blocks (SOF1–SOF3, SOF6) rendered in acidic yellow and lime-green gradients, alongside jagged particle systems (SOF4–SOF5) collectively consistent with crossmodal visual–taste research on sourness (see Table 7).

3.2.3 Audio.

Audio of "SW". The soundtrack we used for "SW" supports a sweet perception (Table 8). High-pitched sounds and piano timbres are associated with sweet tastes [85]. The music features consonant, high-pitched piano notes and bells, aiming to reinforce sweetness perception while complementing the visual environment through soft, flowing melodies.

Figure 4: "SO" elements

Table 2: The ambient elements that are sweet-associated (SWA).

Visual	Element ID	Dimension	Description	Design Rationale
	SWA1	Color	Blooming post-processing effect applied across objects, giving a soft pink glow	Pink bloom builds a sweet-associated atmosphere based on color–taste mappings between pink/red and sweetness [18, 30, 67, 69].
	SWA2	Color	Directional light casting pink-toned illumination across the scene	Warm pink lighting supports sweetness expectations from crossmodal color–taste studies [18, 30, 67, 69].
	SWA3	Shape	Curved surrounding landscape visible from all viewpoints	Use of curved geometry follows findings that rounded and curved shapes are associated with sweet tastes [18, 24, 67].
	SWA3	Color	Landscape feature pink gradients	Pink gradients reinforce a sweet atmosphere via pink–sweet crossmodal associations [18, 30, 67, 69].

Table 3: Elements classified as both ambient and focal that are sweet-associated (SWAF).

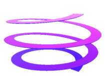
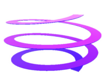
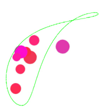
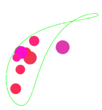
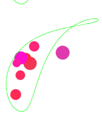
Visual	Element ID	Dimension	Description	Design Rationale
	SWAF1	Shape	Large spiral pathway with a curved profile	Spiral, rounded path strengthens the link between curved forms and sweetness while guiding attention through space [18, 24, 67].
	SWAF1	Color	Pathway rendered with pink gradients	Pink tones maintain a coherent sweet visual theme across the main structure [18, 30, 67, 69].
	SWAF2	Shape	Small circular decorative objects placed along SWAF1	Curved shapes add sweet cues in line with round–sweet correspondences [18, 24, 67].
	SWAF2	Color	Decorative geometry rendered with pink shaders	Red–pink palette supports sweet expectations [18, 30, 67, 69].

Table 4: The focal elements that are sweet-associated (SWF).

Visual	Element ID	Dimension	Description	Design Rationale
	SWF1	Shape	Spheres of varying sizes	Spherical form conforms to roundness associated with sweetness [18, 24, 67].
	SWF1	Motion	Some spheres moving slowly along circular trajectories	Slow, smooth motion is associated with sweetness [32, 38, 67], and circular trajectories resonate with the roundness typically linked to sweet taste [18, 24, 67].
	SWF1	Color	Spheres rendered with the red or pink shader	Red/pink coloration aligns with sweet-linked hues in crossmodal research [18, 30, 67, 69].
	SWF2	Shape	Lollipop-like objects with spherical tops and thin stems	Recognizable lollipop shape is designed to cue remembered experiences of sweet candy, since contextual cues can support the retrieval of associated memories [29].
	SWF2	Color	Lollipop forms shaded with pink gradients shader	Pink color amplifies sweetness expectations through color–taste correspondence [18, 30, 67, 69].
	SWF3	Shape	Circular particle systems	Circular particle layouts avoid sharpness and remain consistent with round–sweet mapping.
	SWF3	Motion	Particles moving in circular-like paths	Circular paths visually fit the roundness association linked to sweetness [18, 24, 67].
	SWF3	Color	The particles were set to transition from light pink to deep pink using the "Start Color" and "Color Over Lifetime" settings.	Pink color range keeps the dynamic effects within the sweet palette [18, 30, 67, 69].
	SWF4	Motion	Particles gently fading in and out	Smooth, slow dynamics are chosen to match sweet-associated motion qualities [32, 38, 67].
	SWF4	Color	The particles were set to transition from light pink to deep pink using the "Start Color" and "Color Over Lifetime" settings.	Pink color range keeps the dynamic effects within the sweet palette [18, 30, 67, 69].

Audio of "SO". The soundtrack we used for "SO" (Table 8) mirrors the sharp, jarring qualities of sourness [85]. The soundtrack features high-pitched tones, staccato rhythms, and occasional dissonant textures.

3.2.4 Environment trajectories. Given that sweetness is the most salient sensory characteristic of sugar-sweetened beverages [51], we believe that a sweet-themed environment serves as an appropriate baseline for adaptation. In this system, we retained the overall

Table 5: The ambient element that is sour-associated (SOA).

Visual	Element ID	Dimension	Description	Design Rationale
	SOA1	Color	Directional light bathing the environment in vivid yellow	Yellow brightness is linked to sour flavor perception in crossmodal studies [67, 69, 83].

Table 6: Elements classified as both ambient and focal that are sour-associated (SOAF).

Visual	Element ID	Dimension	Description	Design Rationale
	SOAF1	Motion	VFX stripes extending across space	Fast motion associated with sour tastes [32, 67].
	SOAF1	Color	Stripes rendered in lighter yellow tones	Yellow hues maintain the sour palette [67, 69, 83].

spatial structure and visual language of the "SW" elements, including ambient elements such as the landscape (SWA3) and the spiral path (SWAF1, SWAF2). These large-scale structures also establish a sense of journey within the virtual environment, guiding participants through a flavor-driven experience.

Upon entering "Flavor Drift" (Fig. 5A), participants encounter SWAF1 and can choose to ascend the spiral pathway ahead from entrance (Fig. 5B) using the thumbstick of the left controller. During the ascent (Fig. 5C-5D), they are able to view different parts of the scene from multiple perspectives, revealing new visual elements as they move upward. While we designed this guided trajectory in a way that we believed the participant might enjoy engaging with it, participants may alternatively choose free exploration. For example, they could visit the hillside area (Fig. 5E) to observe the environment from an elevated perspective, gaining different views of the elements. This open exploration mirrors the flexible, self-directed nature of bubble tea consumption, we hoped.

3.3 Parameterization

The initial environment (Fig. 5) is dominated by SW elements, with SO elements interwoven throughout. Next, we articulate the baseline parameters of the initial environment and the types of changes it supports. We introduced parameterized sensory elements that could be adjusted in real time. Visual forms from both "SW" and "SO" were combined into hybrid, adjustable components, allowing the environment to shift dynamically in response to participant input.

All parameterization details are summarized in Table 9 (Appendix B), which provides a clear overview of each parameter, its related elements, baseline setting, adaptive options, potential triggers (participant input examples), and the resulting sweet–sour effects.

3.3.1 Bloom. The first adjustable parameter is bloom intensity (SWA1)(Fig. 6), a post-processing effect initially set to intensity = 1. In Flavor Drift, we implemented a four-level scale (0–3) to enable fine-grained modulation of ambient glow. Higher bloom levels produce a more radiant, diffused lighting effect, which could be associated with "cozy" or "soft" moods, supporting a more sweetness-congruent atmosphere. Intensity level can be changed dynamically based on the participant's verbal input, reflecting their moment-to-moment flavor preferences.

3.3.2 Color. We also introduce color (Fig. 7) as an adjustable parameter, enabling visual transformations that gave rise to a set of fusion visual elements—hybrid forms blending characteristics from both "SW" and "SO," labeled with an "M" for merge. M1 is a color-adjusted version of SWF3, where the original rounded particle forms from "SW" are recolored in yellow, incorporating the hue typically associated with sourness. M2 repurposes SOF4 by retaining its spiky, fast-moving motion pattern but shifting its color to pink, which reduces its perceived sourness while adding sweet-congruent visual cues. Similarly, M3 and M4 are the yellow-tinted versions of SWAF1 and SWAF2, respectively, allowing these curved, sweet-associated structures to also convey a sour quality through color adaptation. Color changes are triggered based on participants' verbal input of flavor preferences. For example, when a participant indicates a desire for a more sour experience, elements like SWAF1 dynamically may shift to M3 to visually reflect that preference.

3.3.3 Light. Lighting color (Fig. 8) also serves as a parameter. The system includes both SOA1 (Fig. 8A) and SWA2 (Fig. 8B) as lighting presets, though only one can be active at a time according to participant's needs. Switching between them produced noticeable shifts in global ambience, affecting the entire visual tone of the environment as well as the virtual drink. For example, under yellow lighting (SOA1), the same beverage appears more citrus-toned, subtly reinforcing sourness. M5 is based on SWA5, which appears visually yellow-toned under the influence of sour-associated lighting (SOA1). While the object's material properties remain unchanged, its appearance shifts due to the environmental lighting, resulting

Table 7: The focal elements that are sour-associated (SOF).

Visual	Element ID	Dimension	Description	Design Rationale
	SOF1/SOF2/ SOF3/SOF6	Shape	Blocks with harsh, well-defined edges	Geometric forms highlight the sharp, angular character that are associated with sour tastes [26, 32, 67, 75].
	SOF1/SOF2/ SOF3/SOF6	Color	Blocks colored with yellow or lime gradients	Yellow–green tones visually echo sour flavor qualities [67].
	SOF4	Motion	Particles erupting quickly in small explosions	Sudden, fast motion corresponds to the intense, sharp onset of sour tastes [32, 67].
	SOF4	Color	Yellow particles	Bright yellow colors emphasize the sour character [67, 69, 83].
	SOF5	Shape	Flame-like bursts composed of spikes	Spiky geometry associates sour flavor [26, 32, 67, 75].
	SOF5	Motion	Fast flame-mimicking motion	Fast-speed motion associates with sourness [32, 67].
	SOF5	Color	Flame bursts rendered in luminous yellow hues	Luminous yellow colors maintain consistency with sour-linked tones [67, 69, 83].

Table 8: Audio Design

Audio Type	Description	Design Rationale
SW Audio	High-pitched piano & bell-based melodic track	High-pitched sounds and piano timbres are associated with sweet tastes [85].
SO Audio	High-pitched, staccato, slightly dissonant textures	Sharp rhythms and dissonance amplify sourness-linked sensory cues [85].

in a perceptual fusion effect. For this reason, we labeled it as M5. In the initial environment, we used SWA2 as the primary directional lighting, casting a soft pink hue that reinforced the sweet theme and set the baseline atmosphere for subsequent adaptations.

3.3.4 Environmental dynamic effect. Environmental motion dynamics (Fig. 9) are selectively activated. Either the slow-floating particles (SWF4) or the fast-traveling VFX stripes (SOAF1) could be introduced, depending on real-time feedback. These effects remain inactive by default and are triggered only when participants express a need.

3.3.5 Geometry. We also manipulated the number of geometric elements (Fig. 10). Increasing quantity expands the overall visual area occupied by a given category, which amplifies its perceptual salience, we hoped. The initial environment includes more geometry elements from "SW" (Fig. 10B), but the system dynamically increases the quantity of geometry from "SO" (Fig. 10A) if participants request more intensity or sourness. In contrast, if participants express a desire for calm, soft, or sweet tones, the system increases quantity of geometry from "SW" (Fig. 10B).

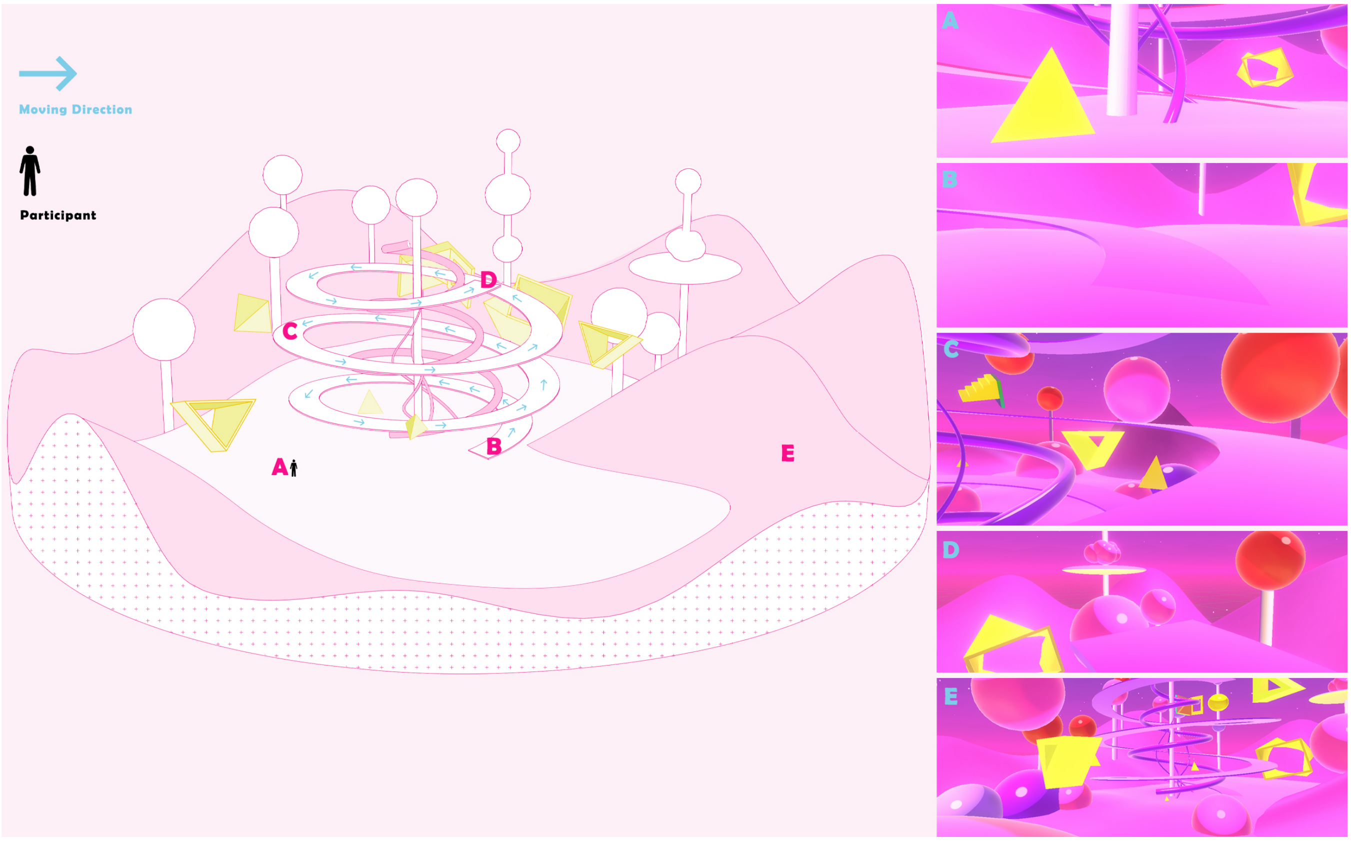


Figure 5: A bird's-eye view illustration showing the trajectories of the "Flavor Drift" virtual environment: A) start location with a human figure, scaled to match the proportions of the environment, used to represent the participant in the virtual environment; B) entrance to the spiral pathway; C) the view while climbing the spiral pathway; D) the end of the spiral pathway; E) free exploration area.

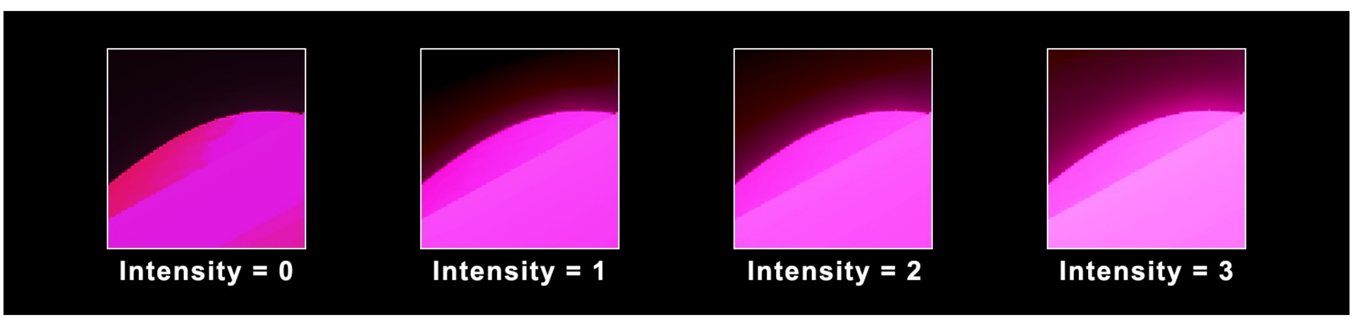


Figure 6: Bloom intensity levels (SWA1) in the "Flavor Drift" environment.

3.3.6 Music. Lastly, we treated music as a dynamic layer. The initial environment begins without background music. Based on the content of the verbal input, the system selects either the sweet-congruent or sour-congruent soundtrack. These auditory elements switch mid-experience in response to shifts in participants input.

Together, these parameterizations enable "Flavor Drift" to function as an adaptive environment that is continuously reshaped by participants' input, though additional parameters may be incorporated in future iterations.

3.4 Real-time adaptation

We integrated GPT-4.5-preview into "Flavor Drift" as an environment adaptation assistant of participant input to enable real-time environmental adaptation across parameters such as color, geometry, and music (Fig. 11).

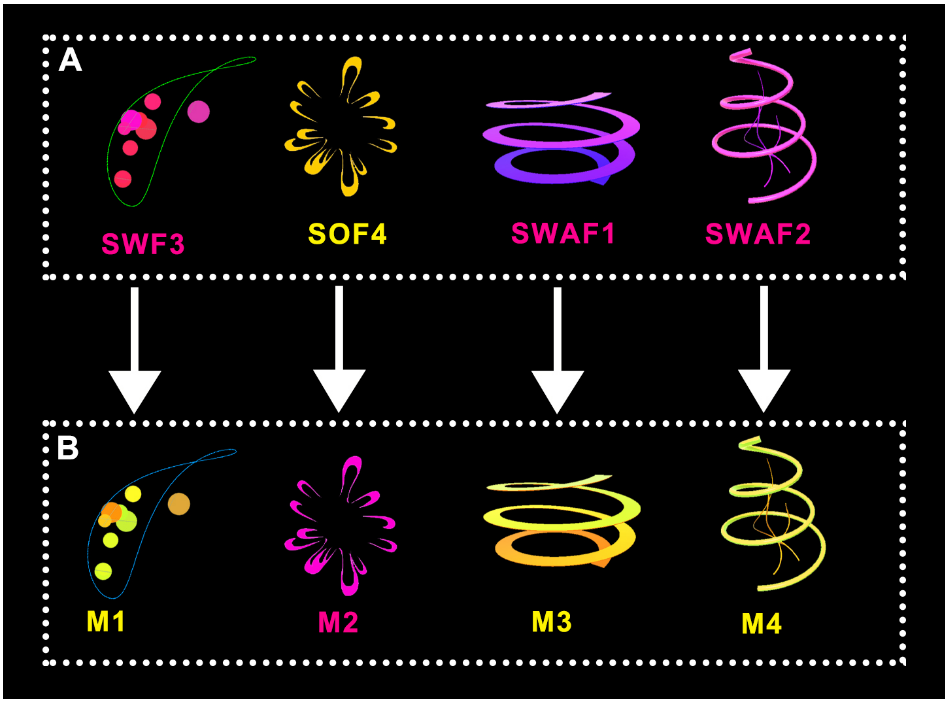


Figure 7: Dynamic color modulation.

To guide GPT-4.5-preview, we adopted a few-shot fine tuning method [34]: the initial prompt included instructions and example input-output pairs that helped the model understand how to extract and translate participant input. Our prompt was designed to encourage step-by-step reasoning, enabling GPT-4.5-preview to identify sensory cues and intended changes, and to return structured commands that accurately map participant input (Appendix A).

The GPT-4.5-preview model's output is a structured set of parameter updates, which is parsed by Unity and applied instantly to adjust the corresponding environmental parameters; these changes, in turn, influence the participant's sensory experience, prompting further input based on their updated perception. We aimed to enable intuitive, uninterrupted interaction by closing the loop between user perception and environmental response.

4 A STUDY ON EXPERIENCING XTEA

To better understand how "XTea" could enrich the bubble tea drinking experience, we conducted a qualitative study. The study received approval from our institution's ethics review board.

4.1 Participants

Twelve participants (aged 19–36 years, M = 26.7, SD = 4.3) were recruited from the local university community. 5 participants identify as female, 7 as male, and none as non-binary. To minimize cultural bias in interoceptive reporting during data collection [76], we ensured the participant pool included individuals from a range of national backgrounds. All participants were self-identified bubble tea enthusiasts [57] and either reported no known allergies or selected allergen-free bubble tea variants based on their dietary needs. Participants were required to have normal or corrected-to-normal vision and no reported impairments in taste perception. Participants were also required to be comfortable wearing a VR headset and engaging in voice-based interaction.

4.2 Procedure

Each study session was conducted in a quiet indoor setting, where one researcher interacted individually with a participant according to a structured sequence. For safety, participants remained seated. The sessions lasted approximately 65–85 minutes and were facilitated by a single researcher. The procedure consisted of four phases: beverage preparation, tutorial session, adaptive experience, and post-experience interview.

4.2.1 Beverage preparation. One day prior to the scheduled session, we contacted each participant to collect their personal bubble tea preferences and make the necessary preparations. A digital menu from a popular local bubble tea shop near the study site was provided, allowing participants to select their preferred flavor, toppings, sugar level, ice level, and choice between hot or cold formats. Participants were also encouraged to request items beyond the standard menu if they had specific preferences; in such cases, the researcher made every reasonable effort to accommodate their requests. The goal was to allow participants to drink a beverage they enjoyed, making the virtual component the only new element compared to their usual consumption experience. The researcher responsible for preparing the drinks had completed a certified Food Safety Level 2 course to ensure hygiene and participant safety. Upon arrival at the study site, participants were welcomed and provided

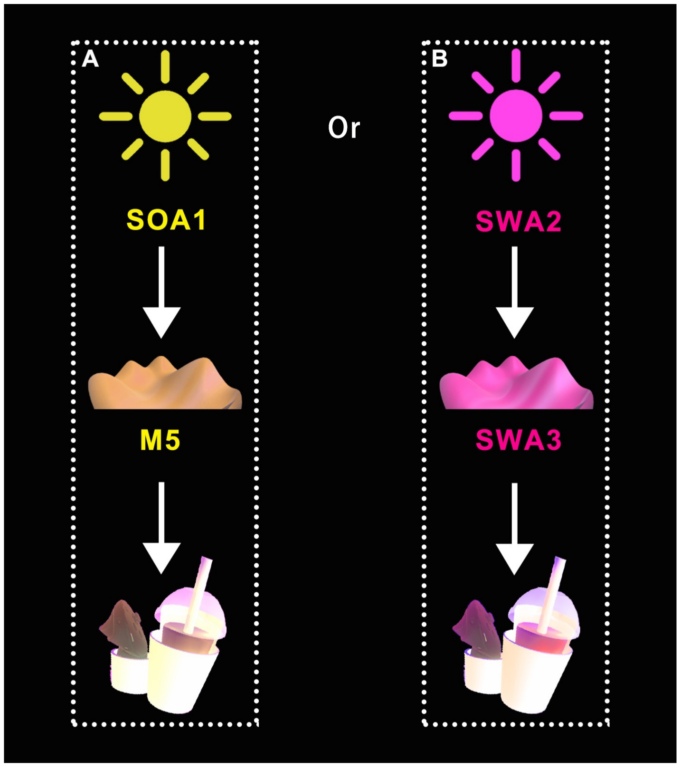


Figure 8: Lighting color as a modulatory parameter for ambient tone and perceptual fusion.

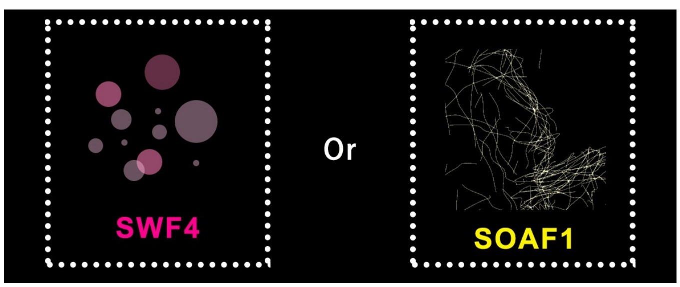


Figure 9: Environmental dynamic effects triggered by participant input.

with an explanatory statement detailing the study's purpose, procedures, and potential risks (approximately 10 minutes). Written informed consent was obtained. Participants then completed a brief eligibility screening form (confirming no known food allergies or sensory impairments). Each participant was subsequently given the personalized drink they had selected prior to the session.

4.2.2 Tutorial session. For participants without prior MR experience (n=2), a tutorial scene was provided to help them become familiar with the virtual environment. Those with previous MR experience were allowed to skip this scene and instead received a brief verbal explanation from the researcher on how to operate the system. Participants wore a head-mounted VR display and held an "XTea" cup while engaging with a non-adaptive version of the "XTea" experience, which lasted approximately 15 minutes. The goal of this session was to ensure participants felt comfortable navigating the system and to minimize confusion during the main study phases. Participants were also encouraged to speak freely during the experience in order to become accustomed to verbal interaction. The virtual environment in this phase remained static and did not respond to user input.

4.2.3 Adaptive experience. Participants then experienced the adaptive version of the system, still drinking the same beverage (approx. 15 minutes). As before, participants were encouraged to speak naturally throughout the experience. During the experience, participants drink bubble tea without time restrictions or pace requirements. The session ends when participants either finish their drink or choose to stop due to fatigue—such as visual discomfort, or general tiredness from extended headset use. Video was recorded throughout.

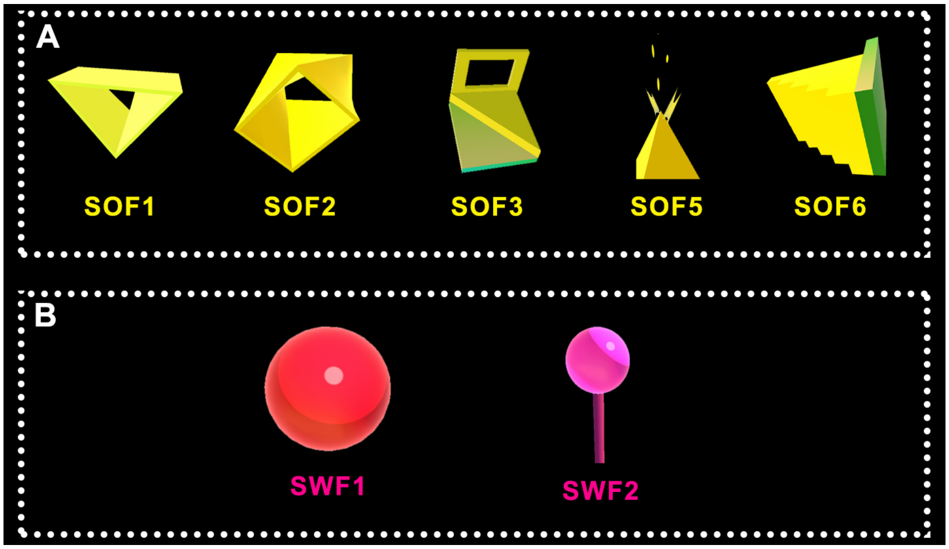


Figure 10: Geometry-based modulation.

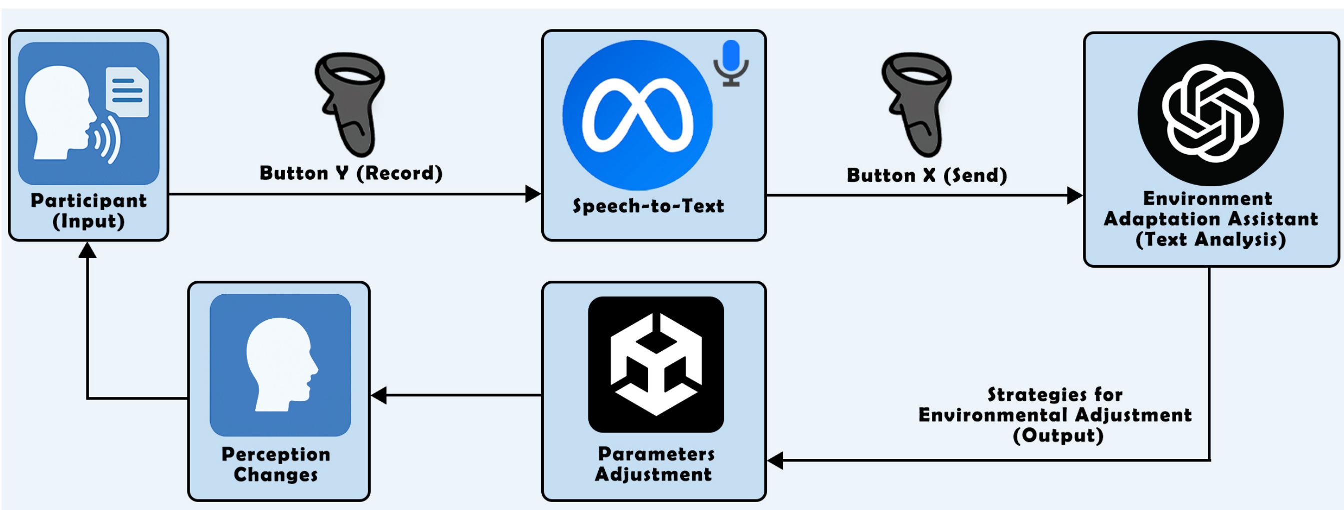


Figure 11: System architecture of XTea

4.2.4 Post-study interview. The participant was afterwards invited to a semi-structured interview (approx. 30 minutes). The micro-phenomenology interview technique [52] was applied to capture subjective experience. The emphasis lied on experiential dimensions such as thought, perception, and sensation [49]. Interviews were audio-recorded and later transcribed for analysis. Participants could skip questions or terminate the interview at any time. Finally, participants were thanked for their participation.

4.3 Data analysis

We conducted an inductive thematic analysis [9, 70], supported by NVivo 15. A master NVivo project was used to facilitate two iterative rounds of open coding by two independent researchers. We treated each question-and-answer pair as a single unit and applied one code to each unit. In this study, saturation was considered to have been reached after ten interviews; two additional interviews were analysed to confirm that no further themes could be identified.

In the initial round, the first coder generated 156 codes and the second 189 codes. After discussing discrepancies, the coders refined the coding schema collaboratively and consolidated it into the master file (270 codes).

A second round of coding was then performed using the updated coding schema, leading to 91 final codes. The coders engaged in multiple rounds of discussion to cluster these codes into higher-order conceptual categories. Through Miro board grouping, six themes were generated. This paper focuses on three of those themes that directly relate to the sensory awareness of participants, the personalized drinking experience, and the perceived agency throughout the adaptive experience.

To complement the thematic analysis, we conducted a parallel review of video recordings to observe physical and behavioral indicators, such as sipping patterns, head and hand movements, and verbal-emotional tone shifts. While these behaviors were not formally coded, they informed our interpretation of attention, bodily engagement, and system responsiveness.

5 FINDINGS

This section presents our findings (F), presented through three themes:

5.1 Theme 1: Enriching sensory engagement through dynamic changes in the virtual environment

Changes in the virtual environment were frequently described as memorable peaks that made the flavor and drinking experience stand out. For example, P8 described the virtual environment as *"surprising"* because *"that's dynamic, that's changing."* P5 explicitly pointed out that changes in the elements of the environment altered the perception of taste, noting, *"When it changed back from the pink to the yellow, I felt I wanted the pink one back. And when it did, it changed how I felt the drink tasted."* These shifts often created lasting impressions and became focal points of the experience.

5.1.1 F1. Salient visual changes intensify sensory engagement. Participants generally found large, noticeable changes in visual features, particularly color, more effective in enhancing sensory engagement (P1, P3, P9, P10). For example, P10 recalled: *"The most memorable moment was when it suddenly turned completely yellow. I felt as if a burst of bubbles suddenly came up in my mouth, just seeing it made me automatically think, 'This is so sour.'"* When asked if the participant had noticed more subtle details, P10 mentioned the small bubbles (SWF3), however, felt they were visually overshadowed by the dominant spiral structure (SWAF1): *"Although I did notice the change in the [SWF3] bubbles, it didn't have as much of a cognitive impact as the change in the spiral structure [SWAF1]."* P10 explained that while all the flavor elements were present from the beginning, elements change amplified specific taste qualities. For example, sweetness might become more prominent, but environmental shifts could heighten their ability to detect even subtle sour notes: *"Even if only 10% of the drink is sour, you can become more sensitive to detecting it."*

Not all participants welcomed large, dramatic changes. P12 initially found them overwhelming, describing them as *"too many things around."* Later, P12 expressed a preference for smaller modifications: *"The sounds and color stayed the same, but just added more objects to the environment [. . .] feeling nice."*

5.1.2 F2. Subtle changes undermine the sense of control. Subtle environmental changes often failed to produce the intended effect. Observations from the researcher revealed that many participants (n=10) did not notice these changes. In one case, P11 explained that the absence of perceived change reduced their sense of control: "*Was there a change at that time? [. . .] The system didn't seem to respond, so I felt the experience was not being led by me."* During this moment, the researcher confirmed from the cast display that a light change had occurred (from SWA2 to SOA1).

5.1.3 F3. Transformations in the virtual drink reshape taste perception. Beyond environmental adjustments, transformations in the virtual drink itself, when noticed, could substantially influence taste perception. The virtual drink (Fig. 2D), which was a bubble tea cup model paired with the "XTea" Cup, changed color in response to environmental lighting (SWA2, SOA1). P10 was pleasantly surprised to detect this: *"I also noticed the drink bottle — the liquid inside would change color along with the scene. For me, it really felt like my perception could be altered a bit because of that."*

However, participants also noticed inconsistencies between the physical drink and its virtual counterpart. P7, P10 and P12 pointed out that changes in the real bubble tea's portion size were not mirrored in the virtual model. As P7 remarked: *"I couldn't see how much I was drinking, I had less control over how quickly I drank it."* Participants suggested that synchronizing these changes could better support mindful drinking.

5.2 Theme 2: Supporting personalized drinking through dynamic agency in the virtual environments

This theme highlights how "XTea" empowered participants to shape their own drinking experience by giving them active control over the sensory environment. Through voice-based interaction, participants were not passive observers but co-creators of their flavor journey. This heightened sense of agency and personalization transformed bubble tea consumption into a responsive and expressive process.

5.2.1 F4. Verbal input activates expressive agency. P7 described the experience as one in which their voice became a new form of interaction: *"a controller"* for a sensory game. The act of speaking triggered immediate changes in the virtual environment, making the system feel alive, as P6 reflected, *"It felt like the system was listening to me, not just showing me something."* In this way, the system encouraged a dialogic relationship between user and environment. This dynamic communication made participants feel responsible for their own multisensory experience, as P10 remarked: *"The scene wouldn't change unless I said something [...] gave me a sense of dominance over the experience."*

5.2.2 F5. Personalization enhances enjoyment. 5 participants appreciated that the system responded not only to their actions but also to their individual preferences. For example, P10 noted: *"It responded to my needs, adjusting the environment based on what I was tasting. I thought that was really great."* P12 echoed this sentiment: *"Every time I've given an input, I could change it to what I wanted to feel."* This quote shows the capacity to turn the environment into a personalized flavor space.

The connection between environmental congruence and enjoyment was especially clear in P1's reflection: *"It matched the state I wanted to be in while drinking bubble tea [that's what] made it more enjoyable."*

Personalization here was not merely about surface-level preference; it created an alignment between internal sensation and external stimuli. As P6 put it, *"Based on my taste, these things [elements] are changing and trying to give me an awesome feeling like that."*

5.3 Theme 3: Cultivating purpose and social presence through digital cues

This theme explores how narrative cues, mission-like trajectories, and subtle social presence cues embedded in the digital experience transformed drinking bubble tea.

5.3.1 F6. Guided drinking creates a sense of direction. 9 participants felt their experience was directed. Movement paths and voice instructions created a sense of *"mission"* (P4). P4 described the act of climbing (SWAF1) in virtual environment as *"a goal to work towards while drinking,"* giving the moment a structure that extended beyond taste. For P4, the voice prompts were not only instructive, but sometimes perceived as companions: *"That sound is like a buddy, help me get the goal"* (P4). However, participants also reflected critically on the guiding voice. For instance, P9 described the voice as more of a passive monitor than a true companion, and imagined that a persistent visual presence like *"a small robot"* might reinforce the quality of the guidance and *"make the companionship feel more tangible"*.

5.3.2 F7. Guiding voices shape the system's social presence. For some participants, the system was not just functional, it had a personality. For example, P7 thought the system guiding voice message took on a human-like role that contextualized the entire experience, stating, *"When I think of my relationship to the system, I think of the voice, the person first, and the environment is just created by the person."* Similarly, P1 compared the system to a chef in an omakase restaurant because she did not know what the chef would serve: *"A bit like omakase"*, suggesting that the evolving environment could be experienced as a curated sensory meal. P12 felt that the system adjusted the environment to match preferences based on verbal input, which led P12 to describe it as *"someone close to you"*: *"the system knows me [. . .] so like a close friend."* Others, like P9, pointed out the lack of conversational depth, viewing the system more like a tool than a social entity: *"It's [a system] just something I use to trigger changes. If I were talking to a person, I'd use more natural language."*

5.3.3 F8. Narrative prompts foster drinking awareness. The narrative voice also encouraged reflective, in-the-moment awareness. System voice messages such as *"Please take a sip of your drink now and enjoy the moment"* were seen as gentle reminders to tune into the sensory experience. P9 described these moments as grounding: *"Even though I was in a study, the system reminded me to enjoy the moment [. . .] whether it was sweet or sour, it could all be enjoyable."* For some, the prompt even acted as a behavioral cue. P2 shared, *"If the [system voice] didn't say it, I wouldn't drink. But if he [system voice] said that, I would."* Through this, the system appeared to reshape the act of drinking as a conscious guided experience.

6 DISCUSSION

Our themes suggest how interaction, agency, and context in a dynamic immersive beverage experience can shape flavor perception. Traditionally, flavor is treated as a fixed property of the drink—unchanging once formulated. In contrast, our system highlights how participants' moment-to-moment interactions with "XTea" suggest possibilities for experiencing flavor as a dynamic, co-created sensory experience. This shift should be understood as an exploratory direction that opens potential avenues for designing, perceiving, and personalizing flavor.

In this section, we explore how our findings extend prior work on computational gastronomy, HFI, and embodied sensory design. We distill our insights into four design strategies to guide interaction designers working with a multisensory immersive experience. Prior work in HFI has focused on offering design strategies (sometimes referred to as design implications) on how to "do" in the future [20, 84]. As HFI is a relatively young field, we encourage exploration rather than only providing strict guidance. Hence, we extend this structure by framing each strategy with a "**Do**," a "**Don't**," and a "**Try**":

Do: Recommended practices grounded in our findings.
Don't: Pitfalls that may hinder a better experience.
Try: Speculative prompts for future exploration.

The "Do/Don't/Try" design strategies we propose could be interpreted as initial design prompts grounded in recurring patterns we observed. Given the exploratory nature of our study, these strategies should be viewed as provisional guidance that may evolve with further empirical work. We also encourage future work to refine, test, or challenge these strategies through more systematic experimentation.

6.1 Supporting sensory engagement with visual and environmental dynamics

Our findings reveal how participants perceived dynamic visual transformations (i.e., environmental changes) as sensory anchors in their flavor experiences (Theme 1). Participants described dynamic changes, particularly in color (P1, P3, P9, P10) and spiral structure (P10), as memorable moments that intensified their engagement and sensation with the drink, suggesting that visually salient dynamic visual cues can more effectively enrich the beverage experience (F1). This result is broadly consistent with prior research on crossmodal correspondences that visual and auditory cues (e.g., color, shape, motion, pitch) can modulate flavor perception [18, 32, 67]. Beyond confirmation, we extend this theory by integrating dynamic digital stimuli into the drinking process. Unlike prior studies that focus on pre-designed or static visual modulation alone [18, 54], our system integrates environmental transformation and drink representation within an interactive virtual environment, allowing dynamic visuals to become an evolving component of the flavor experience. This integration ensures that participants engage in a co-evolving sensory experience, aligning with emerging work in HFI that positions the diner within a dynamic experiential framework [16, 20, 84].

6.1.1 Design Strategy 1: Consider computing salient dynamic visual cues in immersive environment to craft evolving beverage experiences. This strategy builds on two findings from Theme 1 (F1 and F2), which together suggest the possibility that salient, dynamic visual changes could contribute to a more memorable drinking experience (F1) while the limited impact of subtle cues may signal perceptual thresholds in virtual environments (F2). From this, it is plausible that higher-intensity stimuli are better positioned to enhance sensory engagement.

However, subtle changes should not be dismissed entirely. As P10 noted, subtle visual elements (e.g., SWF3) still contributed to the experience. Therefore, subtle stimuli may be retained, but ideally in conjunction with salient transformations, to ensure textural richness in the multisensory experience.

Despite the apparent effectiveness of salient cues, designers should exercise caution. As P12 indicated, a virtual environment that enhances one participant's flavor experience may feel overwhelming to another (F1). This discrepancy underscores individual differences in interoceptive sensibility—the subjective tendency to attend to and interpret bodily sensations [25], such as satiety or flavor intensity, which may influence multisensory experiences. These individual differences imply that immersive flavor systems might need to consider perceptual variability across users. One possible approach is to implement a pre-experience evaluation of users' interoceptive profiles, thereby enabling the system to calibrate stimulation intensities to match each user's unique sensory threshold and preference.

In summary, when designing dynamic immersive drinking experiences, the following prompts may be useful:

- **Do:** Do prioritize salient visual dynamics; allow for layered subtlety; consider interoceptive sensibility.
- **Don't:** Don't rely solely on subtle cues or static configurations.
- **Try:** Try personalization through perceptual profiling that adapts visual intensity.

6.1.2 Design Strategy 2: Consider leveraging dynamic virtual environment to unfold flavors. Dynamic virtual environments have the potential to shift drinking from a passive act into an exploratory experience (Theme 1, Theme3). Findings revealed that visual transitions, particularly those synchronized with the act of drinking, could facilitate heightened awareness of subtle or otherwise overlooked flavor elements (F1, F3). Additionally, the evolving nature of the environment, especially when perceived as unpredictable, may prompt participants to treat the experience as a curated sensory meal (F7). These findings suggest that the temporal dynamics of the virtual environment could serve as an attentional scaffold, potentially enabling the environment to progressively direct participants' awareness toward different aspects of a drink's flavor composition. These findings align with prior work by Obrist et al. [47], which emphasizes that taste is not only detected on the tongue but also perceived as a temporally evolving, multisensory experience. Our study confirms that virtual environment can act as an effective medium for guiding participants' attention through a flavor journey via visual and temporal cues. Beyond simply modifying environmental context, our implementation extends previous multisensory work by introducing the "XTea" cup which allows participants to see the real-time position of their physical cup in virtual environment, and thereby enhancing congruence between drinking actions and sensory feedback. These findings also align with Wang et al.'s work on "SoniCream" [84], wherein digital sound as an additional food layer enriched the playfulness of the ice cream experience. Our study extends this finding to drinking contexts, proposing that dynamic elements in a virtual environment could similarly function as "flavor layers" for beverages. Rather than serving as context or enhancement alone, these dynamic elements might become constitutive parts of the beverage experience itself.

Furthermore, the virtual drink itself emerged as a critical multisensory design element. For example, P10 noted that changes in the color of the virtual drink altered their expectations or interpretations of its taste (F1), reinforcing the importance of aligning visual flavor cues with the virtual drink. However, mismatches were observed, such as the virtual liquid level remaining static despite ongoing physical sipping, suggesting a future direction for refining virtual drink representations to better maintain immersive coherence.

In summary, when designing for flavor unfolding in a virtual environment, the following prompts may be useful:

- **Do:** Do use dynamic visual and auditory transitions in the virtual environment as evolving flavor layers to gradually reveal tastes and encourage playful exploration.
- **Don't:** Don't keep the virtual drink static or let it desynchronize from the user's physical drinking behavior.
- **Try:** Try synchronizing the virtual drink level with physical sipping to maintain immersion and bodily coherence.

6.2 Empowering personalized beverage experience through agency in virtual environments

In HCI, agency refers to the user's ability to act intentionally and exert meaningful control within technological interactions [5]. In the context of immersive flavor experiences, we understand agency as the user's ability to actively personalize the flavor experience through interaction with technological systems. This redefinition of agency is grounded in the dynamic relationship between user input and multisensory output. "XTea" appeared to facilitate a more embodied form of agency by allowing voice-based interaction and therefore altering the sensory cues in real time, not to accomplish a task, but to contribute to create a resonance between internal taste states and external stimuli. This bidirectional tuning process could reframe agency as a kind of sensory creation.

6.2.1 Design Strategy 3: Consider shifting control to the diner by making AI as a sensory assistant. Our findings suggest that enabling dynamic user agency through natural language input could enhance the sense of control of participants during the beverage experience (Theme 2). This dynamic agency was experienced as shifting drinking from a static sensory act toward a more personalized, co-created experience. We extend prior HFI research on computational food experiences, such as "Dancing Delicacies" [21], which emphasized the choreography between food, chef, and diner by allowing chefs to accommodate trajectories of food and customer preferences. While "Dancing Delicacies" enabled diners to engage with pre-scripted dynamic food interactions authored by chefs, control largely remained on the creator's side. "XTea" may extend this by suggesting a shift toward diner-centric environments, inviting participants to shape their own flavor journeys through their own language, and, in doing so, challenging the traditional chef (drink maker)-as-director paradigm.

In addition, "XTea" not only responds to participants but also synchronizes with their taste experiences, potentially empowering them to actively shape their drinking experience by adjusting the sensory environment (Theme 2). This positions XTea as a sensory assistant, extending discussions of AI as an agent [4, 55, 71]. While prior work such as D-Twins [36] has uncovered the potential of AI-driven systems as affective agents—systems that reflect and align with users' emotional states in real time—our work demonstrates the potential of AI-driven systems as sensory agents, extending human sensory perception and agency.

In summary, when designing for diner-driven sensory agency in multisensory beverage experiences, the following prompts may be useful:

- **Do:** Do enable diners to shape their own sensory trajectories through natural language interaction for deeper engagement and personalization.
- **Don't:** Don't restrict sensory experiences to designer-authored scripts that limit user agency.
- **Try:** Try treating AI not just as a conversational agent but as a sensory agent that extends user perceptual experiences in real time.

6.3 Framing social presence in virtual beverage experiences

Our findings reveal that narrative framing and subtle social cues in virtual environment could cultivate a heightened sense of purpose, presence, and reflection during the drinking experience (Theme 3). Participants described how following a guided immersive journey, such as climbing SWAF1, made the act of drinking feel goal-oriented and structured (F6). The system's voice cues were often anthropomorphized, which created a sense of shared presence within the virtual environment (F7). In some cases, the system's prompts encouraged participants to pay attention to their sensations (F8). These responses suggest that narrative trajectories could transform drinking from a passive task into a socially enriched experience. These responses are especially notable given that immersive experiences are often perceived as socially isolating [63]. The emergence of relational and purposeful cues within "XTea" suggests that multisensory systems have potential to counteract immersive environment's isolating tendencies by simulating companionship, transforming solitary acts into guided, socially enriched experience.

6.3.1 Design Strategy 4: Consider designing virtual companions to support social dimensions of drinking. The sense of companionship provided by "XTea" also revealed an reframing: drinking, even when done alone, could be inherently social (Theme 3). Participants like P4 explicitly noted feeling "accompanied" by the system, suggesting that the presence of a virtual companion fulfilled a relational role. However, P9 noted that when the system was limited to voice alone without a visible embodiment, the sense of social presence was diminished, highlighting the potential of visual representation in virtual companionship.

Participants described the system as having a distinct personality, even anthropomorphizing it as a *"chef,"* (P1) a *"friend,"* (P12) or a *"companion,"* (P4). This recurring attribution aligns with broader technological shifts, specifically the growing popularity of AI personas with emotionally resonant identities (e.g., "Monday" in GPT-based applications [4], and Tolan Application[5]), where users increasingly treat computational agents not merely as tools but as social actors [56]. "XTea" extends this emerging dynamic into the context of everyday consumption, demonstrating that personality-infused systems could meaningfully support the social and emotional dimensions of daily rituals like drinking, especially when real-world companionship is absent or inaccessible.

In summary, when designing for socially enriched drinking experiences, the following prompts may be useful:

- **Do:** Do design personality-infused agents that simulate social presence.
- **Don't:** Don't rely solely on visual immersion without offering relational scaffolding.
- **Try:** Try combining voice and visual embodiment to strengthen the sense of companionship.

7 LIMITATIONS AND FUTURE WORK

While "XTea" advances our understanding of how immersive virtual environments could shape flavor perception and engagement, several limitations constrain the current scope of the work and point to compelling directions for future research.

First, the experience studied was situated within a controlled lab setting. Although this setup offered consistency and safety, it may not have fully captured how immersive sensory cues function in more dynamic, real-world environments, such as bubble tea shops, cafés and in social gatherings. Another limitation arises from the constrained ecological validity of current headset-controller-based systems. While bubble tea is typically consumed while walking, commuting, or socializing, our stationary seated setup restricted natural movement and contextual congruence. This discrepancy between our study and naturalistic consumption environments may reduce the generalizability of our findings. Future work should explore how "XTea"-like systems perform outside the lab, and investigate the development of mobile-compatible, multisensory environments that align with habitual drinking behaviors. Such approaches may enhance our understanding of how flavor perception operates in situ, and how adaptive systems could influence everyday food experiences beyond novelty.

Second, the flavor dimensions explored in our study were limited to sweetness and sourness, which we selected for their relevance to bubble tea's typical flavor spectrum. However, flavor experiences are often more complex, and could involve bitterness, umami, aroma, texture, and even thermal cues, as suggested by prior research on drinking augmentation [28]. Future research should expand the design of immersive environments to accommodate and influence a broader palette of gustatory sensations, potentially leading to richer and more ecologically valid applications in health, gastronomy, or behavioral nudging. Furthermore, we acknowledge that participants consumed different self-chosen bubble tea drinks during our study on experiencing XTea, which introduced variation in baseline flavor profiles. This heterogeneity limits the interpretability of any differences in perceived flavor. Future work

[4]https://chatgpt.com/g/g-67ec3b4988f8819184c5454e18f5e84b-monday
[5]https://www.tolans.com/

examining how adaptivity influences perceived taste could employ controlled beverage conditions or systematically manipulate specific flavor dimensions.

Third, although participants reported that voice-based personalization enhanced their sense of agency and enjoyment, our current system had limited conversational depth and interaction flexibility. Future work can focus on expanding the system's capacity for richer, more nuanced dialogue, moving beyond simple command-response interactions toward more fluid, emotionally aware conversations. This could involve integrating affective computing to detect tone and sentiment, enabling the system to respond to not only what users say, but how they say it. Additionally, expanding the system's contextual memory and conversational coherence would allow for more sustained, natural exchanges over the course of a beverage session. This could also explore multimodal feedback loops that combine voice input with gesture, gaze, or physiological signals, opening up new avenues for personalized and intuitive flavor adaptation.

Another limitation is that the liquid level in the virtual cup did not dynamically match the actual drink level. Because our focus in this exploratory study was on experiential qualities rather than perceptual accuracy, we used a simplified model. Nonetheless, this mismatch may have influenced how some visual cues were interpreted. Future implementations could integrate weight sensing or fluid-level tracking to maintain tighter congruence between physical and virtual drinking actions.

We also acknowledge that the present study is exploratory, as our primary aim was to surface design possibilities and experiential qualities of adaptive mixed reality drinking experiences. Future work could benefit from more controlled comparisons, quantitative validation of flavour-related effects, a systematic examination of behavioural or physiological measures beyond self-report, as well as a clearer assessment of the LLM's contribution.

Together, these limitations reflect the early-stage nature of our study. As immersive technologies become more accessible and seamlessly integrated into everyday life, future research can address issues of scalability, contextual flexibility, multisensory richness, methodological robustness, and cross-reality coherence to fully realize the potential of immersive beverage experiences.

8 CONCLUSION

We aimed to broaden the scope of diner-centered HFI by investigating how adaptive, immersive environments can modulate flavor perception and engagement during beverage consumption. Through the design of "XTea", we created a virtual environment that combines sensory adaptation with language-based interaction, and studied how participants shaped and interpreted their beverage experiences in response. Our findings revealed three themes situated at the intersection of personalization, sensory engagement, and social presence. Based upon these themes, we proposed four design strategies for interaction designers interested in crafting immersive beverage experiences. As an early-stage exploration, we hope this work lays the foundation for integrating immersive technologies and natural-language–based intent recognition into everyday beverage experiences, thereby expanding not only how we taste, but also how we experience through consumption.

Acknowledgments

We thank all the participants for their help. Florian 'Floyd' Mueller and Yuchen Zheng thank the Australian Research Council, especially DP190102068, DP200102612, and LP210200656.

REFERENCES

[1] Jeanine Ammann, Christina Hartmann, Vega Peterhans, Sandro Ropelato, and Michael Siegrist. 2020. The relationship between disgust sensitivity and behaviour: A virtual reality study on food disgust. *Food Quality and Preference* 80 (2020), 103833. doi:10.1016/j.foodqual.2019.103833

[2] Y. Annepaka and P. Pakray. 2025. Large language models: a survey of their development, capabilities, and applications. *Knowledge and Information Systems* 67 (2025), 2967–3022. doi:10.1007/s10115-024-02310-4

[3] Ronald G. Bangcuyo, Kacey J. Smith, Jamie L. Zumach, Alex M. Pierce, Gretchen A. Guttman, and Christopher T. Simons. 2015. The use of immersive technologies to improve consumer testing: The role of ecological validity, context and engagement in evaluating coffee. *Food Quality and Preference* 41 (2015), 84–95. doi:10.1016/j.foodqual.2014.11.017

[4] Bahar Bateni and Jim Whitehead. 2024. Language-Driven Play: Large Language Models as Game-Playing Agents in Slay the Spire. In *Proceedings of the 19th International Conference on the Foundations of Digital Games* (Worcester, MA, USA) *(FDG '24)*. Association for Computing Machinery, New York, NY, USA, Article 20, 10 pages. doi:10.1145/3649921.3650013

[5] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 375, 18 pages. doi:10.1145/3544548.3580651

[6] Arpit Bhatia, Henning Pohl, Teresa Hirzle, Hasti Seifi, and Kasper Hornbæk. 2024. Using the Visual Language of Comics to Alter Sensations in Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 603, 17 pages. doi:10.1145/3613904.3642351

[7] B Bhavadharini, V Monica, R Anbarasan, and R Mahendran. 2023. Virtual, augmented, and mixed reality as a versatile tool in food consumer behavior evaluation: Recent advances in aroma, taste, and texture incorporation. *Comprehensive Reviews in Food Science and Food Safety* 22, 6 (2023), 4925–4956. doi:10.1111/1541-4337.13248 arXiv:https://ift.onlinelibrary.wiley.com/doi/pdf/10.1111/1541-4337.13248

[8] Mario Bollini, Stefanie Tellex, Tyler Thompson, Nicholas Roy, and Daniela Rus. 2013. *Interpreting and Executing Recipes with a Cooking Robot*. Springer International Publishing, Heidelberg, 481–495. doi:10.1007/978-3-319-00065-7_33

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[10] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '20)*. Curran Associates Inc., Red Hook, NY, USA, Article 159, 25 pages.

[11] Peter Brusilovsky. 2001. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction* 11, 1 (2001), 87–110. doi:10.1023/A:1011143116306

[12] A. Bschaden, A.F. Dörsam, K. Cvetko, T. Kalamala, and N. Stroebele-Benschop. 2020. The impact of lighting and table linen as ambient factors on meal intake and taste perception. *Food Quality and Preference* 79 (2020), 103797. doi:10.1016/j.foodqual.2019.103797

[13] Lucy Bullivant. 2006. *Responsive Environments: Architecture, Art, and Design.* V & A Publications, London.

[14] Llogari Casas, Samantha Hannah, and Kenny Mitchell. 2024. MoodFlow: Orchestrating Conversations with Emotionally Intelligent Avatars in Mixed Reality. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 86–89. doi:10.1109/VRW62533.2024.00021

[15] Qijia Chen, Andrea Bellucci, and Giulio Jacucci. 2024. "I'd rather drink in VRChat": Understanding Drinking in Social Virtual Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 797, 16 pages. doi:10.1145/3613904.3642405

[16] Weijen Chen, Qingyuan Gao, Zheng Hu, Kouta Minamizawa, and Yun Suen Pai. 2025. Living Bento: Heartbeat-Driven Noodles for Enriched Dining Dynamics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*

(CHI '25). Association for Computing Machinery, New York, NY, USA, Article 353, 18 pages. doi:10.1145/3706598.3713108

[17] Weijen Chen, Yang Yang, Kao-Hua Liu, Yun Suen Pai, Junichi Yamaoka, and Kouta Minamizawa. 2024. Cymatics Cup: Shape-Changing Drinks by Leveraging Cymatics. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 861, 19 pages. doi:10.1145/3613904.3642920

[18] Yang Chen, Arya Xinran Huang, Ilona Faber, Guido Makransky, and Federico J. A. Perez-Cueto. 2020. Assessing the Influence of Visual-Taste Congruency on Perceived Sweetness and Product Liking in Immersive VR. *Foods* 9, 4 (2020). doi:10.3390/foods9040465

[19] Patricia Cornelio, Christopher Dawes, Emanuela Maggioni, Francisco Bernardo, Matti Schwalk, Michaela Mai, Steve Pawlizak, Jingxin Zhang, Gabriele Nelles, Nadejda Krasteva, and Marianna Obrist. 2022. *Virtually tasty*: An investigation of the effect of ambient lighting and 3D-shaped taste stimuli on taste perception in virtual reality. *International Journal of Gastronomy and Food Science* 30 (Dec. 2022), 100626. doi:10.1016/j.ijgfs.2022.100626

[20] Jialin Deng, Yinyi Li, Hongyue Wang, Ziqi Fang, and Florian 'Floyd' Mueller. 2025. Sonic Delights: Exploring the Design of Food as An Auditory-Gustatory Interface. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 356, 19 pages. doi:10.1145/3706598.3713892

[21] Jialin Deng, Nathalie Overdevest, Patrick Olivier, and Florian Mueller. 2024. From Plating to Tasting: Towards Understanding the Choreography of Computational Food. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 359, 17 pages. doi:10.1145/3613904.3642516

[22] Ophelia Deroy and Charles Spence. 2016. Crossmodal Correspondences: Four Challenges. *Multisensory Research* 29, 1-3 (2016), 29 – 48. doi:10.1163/22134808-00002488

[23] Paul Dourish. 2001. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA.

[24] M. T. Fairhurst, D. Pritchard, D. Ospina, et al. 2015. Bouba–Kiki in the plate: combining crossmodal correspondences to change flavour experience. *Flavour* 4, 22 (2015). doi:10.1186/s13411-015-0032-2

[25] Sarah N. Garfinkel, Anil K. Seth, Adam B. Barrett, Keisuke Suzuki, and Hugo D. Critchley. 2015. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology* 104 (2015), 65–74. doi:10.1016/j.biopsycho.2014.11.004

[26] Ignacio Gil-Pérez, Rubén Rebollar, Iván Lidón, Javier Martín, Hans C.M. van Trijp, and Betina Piqueras-Fiszman. 2019. Hot or not? Conveying sensory information on food packaging through the spiciness-shape correspondence. *Food Quality and Preference* 71 (2019), 197–208. doi:10.1016/j.foodqual.2018.07.009

[27] Andrea Grimes and Richard Harper. 2008. Celebratory technology: new directions for food research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 467–476. doi:10.1145/1357054.1357130

[28] Yuki Hashimoto, Naohisa Nagaya, Minoru Kojima, Satoru Miyajima, Junichiro Ohtaki, Akio Yamamoto, Tomoyasu Mitani, and Masahiko Inami. 2006. Straw-like user interface: virtual experience of the sensation of drinking using a straw. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Hollywood, California, USA) *(ACE '06)*. Association for Computing Machinery, New York, NY, USA, 42–es. doi:10.1145/1178823.1178873

[29] Rachel S. Herz. 2004. A Naturalistic Analysis of Autobiographical Memories Triggered by Olfactory Visual and Auditory Stimuli. *Chemical Senses* 29, 3 (March 2004), 217–224. doi:10.1093/chemse/bjh025 _eprint: https://academic.oup.com/chemse/article-pdf/29/3/217/915493/bjh025.pdf.

[30] Souta Hidaka and Kazumasa Shimoda. 2014. Investigation of the Effects of Color on Judgments of Sweetness Using a Taste Adaptation Method. *Multisensory Research* 27, 3-4 (2014), 189–205. doi:10.1163/22134808-00002455

[31] Jingyi Huang. 2023. A Study on the Effect of Color on Human Food Perception. *Journal of Education, Humanities and Social Sciences* 10 (04 2023), 225–230. doi:10.54097/ehss.v10i.7021

[32] Gijs Huisman, Merijn Bruijnes, and Dirk K. J. Heylen. 2016. A Moving Feast: Effects of Color, Shape and Animation on Taste Associations and Taste Perceptions. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology* (Osaka, Japan) *(ACE '16)*. Association for Computing Machinery, New York, NY, USA, Article 13, 12 pages. doi:10.1145/3001773.3001776

[33] Istijanto and Indria Handoko. 2021. What approach and avoidance factors drive Gen-Z consumers to buy bubble tea? An exploratory study. *Young Consumers* 23, 3 (Dec. 2021), 382–396. doi:10.1108/YC-08-2021-1376 Publisher: Emerald Publishing Limited.

[34] Rahul Jain, Amit Goel, Koichiro Niinuma, and Aakar Gupta. 2025. AdaptiveSliders: User-aligned Semantic Slider-based Editing of Text-to-Image Model Output. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 541, 27 pages. doi:10.1145/3706598.3714292

[35] Rohit Ashok Khot and Florian Mueller. 2019. Human-Food Interaction. *Foundations and Trends® in Human–Computer Interaction* 12, 4 (2019), 238–415. doi:10.1561/1100000074

[36] I-Chen Lo and Pei-Luen Patrick Rau. 2025. D-Twins: Your Digital Twin Designed for Real-Time Boredom Intervention. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 321, 15 pages. doi:10.1145/3706598.3714163

[37] Yuhan Luo, Young-Ho Kim, Bongshin Lee, Naeemul Hassan, and Eun Kyoung Choe. 2021. FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 606–618. doi:10.1145/3461778.3462074

[38] Bruno Mesz, Marcos A. Trevisan, and Mariano Sigman. 2011. The taste of music. *Perception* 40, 2 (2011), 209–219. doi:10.1068/p6801

[39] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1994. Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies* 2351 (01 1994). doi:10.1117/12.197321

[40] Fnu Mohbat and Mohammed J. Zaki. 2025. KERL: Knowledge-Enhanced Personalized Recipe Recommendation using Large Language Models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Vienna, Austria, 19125–19141. doi:10.18653/v1/2025.acl-long.938

[41] Florian 'Floyd' Mueller, Richard Byrne, Josh Andres, and Rakesh Patibanda. 2018. Experiencing the Body as Play. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173784

[42] Florian 'Floyd' Mueller, Marianna Obrist, Soh Kim, Masahiko Inami, and Jialin Deng. 2023. Eat-IT: Towards Understanding Interactive Technology and Food (Dagstuhl Seminar 22272). *Dagstuhl Reports* 12, 7 (2023), 19–40. doi:10.4230/DagRep.12.7.19

[43] Florian 'Floyd' Mueller, Yan Wang, Zhuying Li, Tuomas Kari, Peter Arnold, Yash Dhanpal Mehta, Jonathan Marquez, and Rohit Ashok Khot. 2020. Towards Experiencing Eating as Play. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '20)*. Association for Computing Machinery, New York, NY, USA, 239–253. doi:10.1145/3374920.3374930

[44] Florian 'Floyd' Mueller, Marianna Obrist, Ferran Altarriba Bertran, Neharika Makam, Soh Kim, Christopher Dawes, Patrizia Marti, Maurizio Mancini, Eleonora Ceccaldi, Nandini Pasumarthy, Sahej Claire, Kyung seo Jung, Jialin Deng, Jürgen Steimle, Nadejda Krasteva, Matti Schwalk, Harald Reiterer, Hongyue Wang, and Yan Wang. 2024. Grand challenges in human-food interaction. *International Journal of Human-Computer Studies* 183 (2024), 103197. doi:10.1016/j.ijhcs.2023.103197

[45] Takuji Narumi, Shinya Nishizaka, Takashi Kajinami, Tomohiro Tanikawa, and Michitaka Hirose. 2011. MetaCookie+. In *2011 IEEE Virtual Reality Conference*. 265–266. doi:10.1109/VR.2011.5759500

[46] Masahiro Nishizawa, Wanting Jiang, and Katsunori Okajima. 2016. Projective-AR system for customizing the appearance and taste of food. In *Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality* (Tokyo, Japan) *(MVAR '16)*. Association for Computing Machinery, New York, NY, USA, Article 6, 6 pages. doi:10.1145/3001959.3001966

[47] Marianna Obrist, Rob Comber, Sriram Subramanian, Betina Piqueras-Fiszman, Carlos Velasco, and Charles Spence. 2014. Temporal, affective, and embodied characteristics of taste experiences: a framework for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2853–2862. doi:10.1145/2556288.2557007

[48] J. Pallasmaa. 2012. *The Eyes of the Skin: Architecture and the Senses*. Wiley. https://books.google.com.au/books?id=VXUxwHx9wlQC

[49] Stephen E. Palmer and Karen B. Schloss. 2010. An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences* 107, 19 (2010), 8877–8882. doi:10.1073/pnas.0906172107 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.0906172107

[50] Kyösti Pennanen, Johanna Närväinen, Saara Vanhatalo, Roope Raisamo, and Nesli Sozer. 2020. Effect of virtual eating environment on consumers' evaluations of healthy and unhealthy snacks. *Food Quality and Preference* 82 (2020), 103871. doi:10.1016/j.foodqual.2020.103871

[51] Barry M. Popkin and Corinna Hawkes. 2016. Sweetening of the global diet, particularly beverages: patterns, trends, and policy responses. *The Lancet Diabetes & Endocrinology* 4, 2 (2016), 174–186. doi:10.1016/S2213-8587(15)00419-2

[52] Mirjana Prpa, Sarah Fdili-Alaoui, Thecla Schiphorst, and Philippe Pasquier. 2020. Articulating Experience: Reflections from Experts Applying Micro-Phenomenology to Design Research in HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376664

[53] Nimesha Ranasinghe, Kuan-Yi Lee, and Ellen Yi-Luen Do. 2014. FunRasa: an interactive drinking platform. In *Proceedings of the 8th International Conference*

on Tangible, Embedded and Embodied Interaction (Munich, Germany) *(TEI ’14)*. Association for Computing Machinery, New York, NY, USA, 133–136. doi:10.1145/2540930.2540939

[54] Nimesha Ranasinghe, Thi Ngoc Tram Nguyen, Yan Liangkun, Lien-Ya Lin, David Tolley, and Ellen Yi-Luen Do. 2017. Vocktail: A Virtual Cocktail for Pairing Digital Taste, Smell, and Color Sensations. In *Proceedings of the 25th ACM International Conference on Multimedia* (Mountain View, California, USA) *(MM ’17)*. Association for Computing Machinery, New York, NY, USA, 1139–1147. doi:10.1145/3123266.3123440

[55] Alan Rychert, María Luján Ganuza, and Matias Nicolás Selzer. 2024. Integrating GPT as an Assistant for Low-Cost Virtual Reality Escape-Room Games. *IEEE Computer Graphics and Applications* 44, 4 (2024), 14–25. doi:10.1109/MCG.2024.3426314

[56] HEIKE SCHAUMBURG. 2001. COMPUTERS AS TOOLS OR AS SOCIAL ACTORS? — THE USERS’ PERSPECTIVE ON ANTHROPOMORPHIC AGENTS. *International Journal of Cooperative Information Systems* 10, 01n02 (2001), 217–234. doi:10.1142/S0218843001000321 arXiv:https://doi.org/10.1142/S0218843001000321

[57] Thecla Schiphorst. 2011. Self-evidence: applying somatic connoisseurship to experience design. In *CHI ’11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA ’11)*. Association for Computing Machinery, New York, NY, USA, 145–160. doi:10.1145/1979742.1979640

[58] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in Human–Agent Interaction: A Survey. *ACM Comput. Surv.* 54, 4, Article 81 (May 2021), 43 pages. doi:10.1145/3386867

[59] Alon Shoa, Ramon Oliva, Mel Slater, and Doron Friedman. 2023. Sushi with Einstein: Enhancing Hybrid Live Events with LLM-Based Virtual Humans. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents* (Würzburg, Germany) *(IVA ’23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 6 pages. doi:10.1145/3570945.3607317

[60] C. A. Silva. 2006. Liquid Architectures: Marcos Novak’s Territory of Information. In *Proceedings of the 10th International Conference on Information Visualisation (IV’06)*. 693–698. doi:10.1109/IV.2006.70

[61] Fiorella Sinesio, Elisabetta Moneta, Christelle Porcherot, Silvia Abbà, Lise Dreyfuss, Kévin Guillamet, Seppe Bruyninckx, Charles Laporte, Sven Henneberg, and Jean A. McEwan. 2019. Do immersive techniques help to capture consumer reality? *Food Quality and Preference* 77 (2019), 123–134. doi:10.1016/j.foodqual.2019.05.004

[62] Richard Skarbez, Missie Smith, and Mary C. Whitton. 2021. Revisiting Milgram and Kishino’s Reality-Virtuality Continuum. *Frontiers in Virtual Reality* Volume 2 - 2021 (2021). doi:10.3389/frvir.2021.647997

[63] Nikolay Slivkin, Leila Elgaaied-Gambier, and Linda Hamdi-Kidar. 2025. Is Virtual Reality Lonely? The VR–Isolation Stereotype and Its Impact on VR Adoption. *Psychology & Marketing* 42, 4 (2025), 1110–1131. doi:10.1002/mar.22167

[64] Charles Spence. 2015. Multisensory Flavor Perception. *Cell* 161, 1 (2015), 24–35. doi:10.1016/j.cell.2015.03.007

[65] C. Spence. 2016. 1 - Multisensory Packaging Design: Color, Shape, Texture, Sound, and Smell. In *Integrating the Packaging and Product Experience in Food and Beverages*, Peter Burgess (Ed.). Woodhead Publishing, 1–22. doi:10.1016/B978-0-08-100356-5.00001-2

[66] Charles Spence. 2023. Digitally enhancing tasting experiences. *International Journal of Gastronomy and Food Science* 32 (June 2023), 100695. doi:10.1016/j.ijgfs.2023.100695

[67] Charles Spence. 2023. Explaining Visual Shape–Taste Crossmodal Correspondences. *Multisensory Research* 36, 4 (2023), 313 – 345. doi:10.1163/22134808-bja10096

[68] Charles Spence and Alberto Gallace. 2011. Multisensory design: Reaching out to touch the consumer. *Psychology & Marketing* 28, 3 (2011), 267–308. doi:10.1002/mar.20392 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.20392.

[69] Charles Spence, Xiaoang Wan, Andy Woods, Carlos Velasco, Jialin Deng, Jozef Youssef, and Ophelia Deroy. 2015. On tasty colours and colourful tastes? Assessing, explaining, and utilizing crossmodal correspondences between colours and basic tastes. *Flavour* 4, 1 (July 2015), 23. doi:10.1186/s13411-015-0033-1

[70] Vicki Squires. 2023. Thematic Analysis. In *Varieties of Qualitative Research Methods: Selected Contextual Perspectives*, Janet Mola Okoko, Scott Tunison, and Keith D. Walker (Eds.). Springer International Publishing, Cham, 463–468. doi:10.1007/978-3-031-04394-9_72

[71] Ian Steenstra, Farnaz Nouraei, Mehdi Arjmand, and Timothy Bickmore. 2024. Virtual Agents for Alcohol Use Counseling: Exploring LLM-Powered Motivational Interviewing. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents* (GLASGOW, United Kingdom) *(IVA ’24)*. Association for Computing Machinery, New York, NY, USA, Article 20, 10 pages. doi:10.1145/3652988.3673932

[72] Lorena Stäger, Marte Roel Lesur, and Bigna Lenggenhager. 2021. What Am I Drinking? Vision Modulates the Perceived Flavor of Drinks, but No Evidence of Flavor Altering Color Perception in a Mixed Reality Paradigm. *Frontiers in Psychology* Volume 12 - 2021 (2021). doi:10.3389/fpsyg.2021.641069

[73] Anna T. Thomas, Adam Yee, Andrew Mayne, Maya B. Mathur, Dan Jurafsky, and Kristina Gligorić. 2025. What can large language models do for sustainable food? arXiv:2503.04734 [cs.CY]

[74] ANNE. M. TREISMAN. 1964. SELECTIVE ATTENTION IN MAN. *British Medical Bulletin* 20, 1 (Jan. 1964), 12–16. doi:10.1093/oxfordjournals.bmb.a070274 _eprint: https://academic.oup.com/bmb/article-pdf/20/1/12/960075/20-1-12.pdf.

[75] N. Turoman, Carlos Velasco, Y. C. Chen, et al. 2018. Symmetry and its role in the crossmodal correspondence between shape and taste. *Attention, Perception, & Psychophysics* 80 (2018), 738–751. doi:10.3758/s13414-017-1463-x Published online: 19 December 2017; Issue date: April 2018.

[76] Deniz G. Ural, Gabriela Aceves Sepúlveda, and Bernhard E. Riecke. 2025. Who Defines Embodiment? Cultural Bias in Interoceptive Wellness Technologies. In *Companion Publication of the 2025 ACM Designing Interactive Systems Conference (DIS ’25 Companion)*. Association for Computing Machinery, New York, NY, USA, 363–368. doi:10.1145/3715668.3736349

[77] Carlos Velasco, Francisco Barbosa Escobar, Olivia Petit, and Qian Janice Wang. 2021. Impossible (Food) Experiences in Extended Reality. *Frontiers in Computer Science* 3 (Aug. 2021). doi:10.3389/fcomp.2021.716846 Publisher: Frontiers.

[78] Carlos Velasco and Marianna Obrist. 2020. *Multisensory Experiences: Where the senses meet technology.* Oxford University Press. doi:10.1093/oso/9780198849629.001.0001 arXiv:https://academic.oup.com/book/31941/book-pdf/50986691/9780192589538_web.pdf

[79] Carlos Velasco, Marianna Obrist, Olivia Petit, and Charles Spence. 2018. Multisensory Technology for Flavor Augmentation: A Mini Review. *Frontiers in Psychology* Volume 9 - 2018 (2018). doi:10.3389/fpsyg.2018.00026

[80] Carlos Velasco and Charles Spence (Eds.). 2019. *Multisensory Packaging: Designing New Product Experiences.* Palgrave Macmillan Cham, Cham. doi:10.1007/978-3-319-94977-2

[81] Carlos Velasco, Qian Janice Wang, Marianna Obrist, and Anton Nijholt. 2021. A Reflection on the State of Multisensory Human–Food Interaction Research. *Frontiers in Computer Science* 3 (Dec. 2021). doi:10.3389/fcomp.2021.694691 Publisher: Frontiers.

[82] Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria, Bingsheng Yao, Dakuo Wang, Yvonne Coady, and Mirjana Prpa. 2024. Building LLM-based AI Agents in Social Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA ’24)*. Association for Computing Machinery, New York, NY, USA, Article 65, 7 pages. doi:10.1145/3613905.3651026

[83] Xiaoang Wan, Andy T. Woods, Jasper J. F. van den Bosch, Kirsten J. McKenzie, Carlos Velasco, and Charles Spence. 2014. Cross-cultural differences in cross-modal correspondences between basic tastes and visual features. *Frontiers in Psychology* 5 (2014), 1365. doi:10.3389/fpsyg.2014.01365

[84] Hongyue Wang, Jialin Deng, Linjia He, Nathalie Overdevest, Ryan Wee, Yan Wang, Phoebe O. Toups Dugas, Don Samitha Elvitigala, and Florian Floyd Mueller. 2025. Towards Understanding Interactive Sonic Gastronomy with Chefs and Diners. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI ’25)*. Association for Computing Machinery, New York, NY, USA, Article 358, 19 pages. doi:10.1145/3706598.3714237

[85] Qian Wang, Andrew Woods, and Charles Spence. 2015. "Whats Your Taste in Music?" A Comparison of the Effectiveness of Various Soundscapes in Evoking Specific Tastes. *i-Perception* 6 (12 2015). doi:10.1177/2041669515622001

[86] Qian Janice Wang, Rachel Meyer, Stuart Waters, and David Zendle. 2020. A Dash of Virtual Milk: Altering Product Color in Virtual Reality Influences Flavor Perception of Cold-Brew Coffee. *Frontiers in Psychology* 11 (Dec. 2020). doi:10.3389/fpsyg.2020.595788 Publisher: Frontiers.

[87] Yan Wang, Zhuying Li, Rohit Ashok Khot, and Florian ’Floyd’ Mueller. 2022. Toward Understanding Playful Beverage-based Gustosonic Experiences. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 33 (March 2022), 23 pages. doi:10.1145/3517228

[88] Brian Wansink and Koert van Ittersum. 2012. Fast Food Restaurant Lighting and Music can Reduce Calorie Intake and Increase Satisfaction. *Psychological Reports* 111, 1 (2012), 228–232. doi:10.2466/01.PR0.111.4.228-232 PMID: 23045865.

[89] Zhongqi Yang, Elahe Khatibi, Nitish Nagesh, Mahyar Abbasian, Iman Azimi, Ramesh Jain, and Amir Rahmani. 2024. ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health* 32 (06 2024), 100465. doi:10.1016/j.smhl.2024.100465

[90] Yuehao Yin, Huiyan Qi, Bin Zhu, Jingjing Chen, Yu-Gang Jiang, and Chong-Wah Ngo. 2023. FoodLMM: A Versatile Food Assistant using Large Multi-modal Model. arXiv:2312.14991 [cs.CV]

[91] Zheyuan Zhang, Zehong Wang, Tianyi Ma, Varun Sameer Taneja, Sofia Nelson, Nhi Ha Lan Le, Keerthiram Murugesan, Mingxuan Ju, Nitesh V. Chawla, Chuxu Zhang, and Yanfang Ye. 2025. MOPI-HFRS: A Multi-objective Personalized Health-aware Food Recommendation System with LLM-enhanced Interpretation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1* (Toronto ON, Canada) *(KDD ’25)*. Association for Computing Machinery, New York, NY, USA, 2860–2871. doi:10.1145/3690624.3709382

[92] Yuchen Zheng, Jialin Deng, Hongyue Wang, Florian ‘Floyd’ Mueller, Xueni Pan, and Marco Fyfe Pietro Gillies. 2025. immersiTea: Exploring Multisensory Virtual Reality Environments to Enrich Bubble Tea Drinking Experiences. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA ’25)*. Association for Computing Machinery, New York, NY, USA,

Article 613, 7 pages. doi:10.1145/3706599.3720228
[93] Süeda Özkaya, Santiago Berrezueta-Guzman, and Stefan Wagner. 2025. How LLMs are Shaping the Future of Virtual Reality. arXiv:2508.00737 [cs.HC] https://arxiv.org/abs/2508.00737

A Prompt to GPT for mapping sensory preferences from participant input

[Role]
You are a real-time environment adaptation assistant for a virtual reality bubble tea-drinking experience called XTea. The user will speak a sentence after drinking. You will be given the user's sentence, and decide how to adjust the environment using these parameters:

1. Post-processing bloom intensity (SWA1): increase for softer feeling, reduce for sharper atmosphere.

2. Music track selection: activate between sweet music or sour music.

3. Directional lighting tone: activate between yellow (SOA1) and pink (SWA2).

4. Large-scale particle system: activate SWF4 (for sweet) or SOAF1 (for sour).

5. Element ratio control: Increase/decrease how many sweet/sour objects are in the scene.

Use SWF1 to represent sweet; use SOF1-SOF2, SOF5-SOF6 for sour.

6. **Color theme of key geometry and particles** — adjust hue of major elements:

- Set ** SWF3 **, ** SOF4**, ** SWAF1**,** SWAF2** to pink for sweetness

- Set ** SWF3 **, ** SOF4**, ** SWAF1**,** SWAF2** to yellow for sourness

[Task]
Your job is to interpret the user's statement and generate a list of adaptation suggestions. Be specific and concrete. Refer directly to elements using their names (e.g., "increase SWF1", "activate SWF4", "set lighting to SWA2").
[Context]
SW represents Sweet. We categorized our SW elements into three types: SWA, SWF, and SWAF...
...
...
[Output Format]
Respond with exactly 4 lines:
1. **Overall mood interpretation** (1 sentence max)
2. **Which elements should be increased or activated**
Follow the format: Activate[SWF4/SOAF1], Increase[SWA1], Increase[SWF1], Increase [SOF].
3. **Which elements should be reduced**
Follow the format: Reduce[SWA1], Reduce[SWF1], Reduce[SOF].
4. **Changes to lighting, color, music**
Follow the format: Light[SOA1/SWA2], SWF3[Pink/Yellow], SOF4[Pink/Yellow], SWAF1[Pink/Yellow], SWAF2[Pink/Yellow], Music[Sweet/Sour].
[Examples]
INPUT:
It feels too sour. I want something sweeter and smoother.
OUTPUT:
1. The environment needs to shift toward sweetness and smoothness.
2. Activate[SWF4], increase[SWA1].
3. Reduce[SOF]
4. Light[SWA2], Music[Sweet], SWAF1[Pink]
...
...
...
[Constraints]
1. Please ensure that your output strictly follows the output format.
2. Do not repeat the user's input.

B Additional Table

Table 9: Overview of adaptive parameters in *Flavor Drift*.

Parameter	Related elements	Baseline setting	Adaptive options	Potential Triggers (participant input example)	Sweet–sour effects
Bloom intensity	SWA1	Bloom level = 1 (0–3 scale); low ambient glow.	Four-level scale (0–3); higher values increase ambient glow and diffusion of highlights in the scene.	When participants describe the drink/scene as "too sour/sharp" or "too sweet/soft" and request a sweeter (softer) or sourer (sharper) feeling.	Higher bloom → softer, cozy, sweet-congruent atmosphere; lower bloom → sharper, less sweet-biased.
Color of fusion elements	M1–M4 (from SWF3, SOF4, SWAF1–2)	Pink-rounded SW elements and yellow-spiky SO elements are clearly distinguishable.	Recolor rounded forms to yellow to increase sourness (e.g., M1, M3, M4); recolor spiky/motion forms to pink to soften sourness(e.g., M2).	Verbal preferences for "more sour", "less sour", or "more sweet" in the overall atmosphere.	Yellow shifts elements toward sourness; pink softens sour forms and adds sweet-congruent cues.
Lighting color	SOA1 (yellow), SWA2 (pink), M5	Global lighting initially uses SWA2 (pink directional light), providing a warm, sweet-congruent ambience.	Switch between pink SWA2 and yellow SOA1 as the dominant light; under SOA1, the object SWA5 appears as M5.	Requests for a "brighter / citrusy" vs. "soft / warm" ambience, or references to sour vs. sweet moods.	Yellow global light reinforces sourness (including the perceived colour of the drink); pink light maintains a sweet, warm, cozy environment.
Environmental dynamic effect	SWF4, SOAF1	No dynamic environmental effects are active by default; the environment is visually static.	Activate either slow-floating SWF4 particles or fast-travelling SOAF1 stripes.	Participants request the scene to feel sweeter or sourer.	Slow motion → calm, sweet-congruent; fast motion → energetic, intense, sour-congruent dynamism.
Geometry quantity	SW geometry, SO geometry	Environment initially contains more geometry from SW, emphasising sweet flavor in bubble tea.	Increase the proportion of SO geometry to heighten intensity and sourness; increase SW geometry to emphasise calm, soft, sweet-congruent qualities.	Participants request "more sour" or "more sweet" flavor.	More SO forms → sharper, sour-congruent; more SW forms → rounder, sweet-congruent environment.
Music layer	Sweet / sour soundtrack	No background music is played at the beginning of the experience.	Add or switch between sweet-congruent and sour-congruent soundtracks while the participant is exploring the environment.	Participants indicate preference for a sweeter vs. more sour atmosphere, or request the drink to "taste sweeter" or "taste more sour".	Sweet soundtrack → taste sweet, feel cozy, pleasant; sour soundtrack → taste sour, energetic, tangy.