

# InBite: Towards Designing Real-time Configurable Auditory Dining Experience for Each Bite

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

Yuting Wu  
Academy of Music  
Hong Kong Baptist University  
Hong Kong, China  
24484369@life.hkbu.edu.hk

Subasinghe Piyarathnage  
Sahan Madusanka  
Exertion Games Lab, Human-Centred  
Computing, Monash University  
Melbourne, Australia  
sahan.subasinghepiyarathnage@monash.edu

Ziqi Fang  
Exertion Games Lab, Human-Centred  
Computing, Monash University  
Melbourne, VIC, Australia  
ziqi.fang@monash.edu

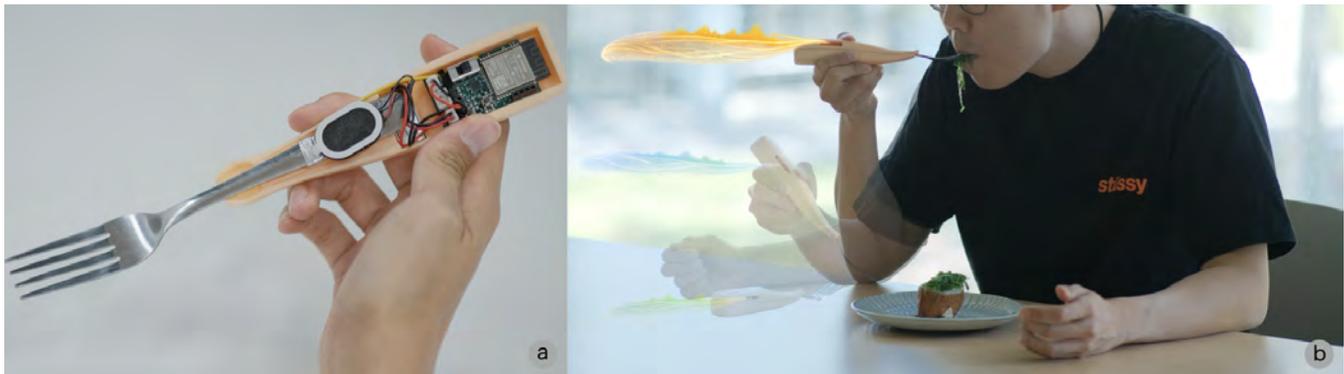
Yuchen Zheng  
Exertion Games Lab, Human-Centred  
Computing, Monash University  
Melbourne, Australia  
yuchen.zheng@monash.edu

Hong Luo  
Exertion Games Lab, Human-Centred  
Computing, Monash University  
Melbourne, Australia  
hong.luo@monash.edu

Jialin Deng  
Diverse-ability Interaction Lab,  
School of Computer Science,  
University of Bristol  
Bristol, United Kingdom  
jialin.deng@bristol.ac.uk

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

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



**Figure 1: Showcasing InBite.** (a) Cutaway view of InBite, showing the custom-designed PCB, a speaker, and a miniature lithium battery. A capacitive-sensing lead extends from the PCB to the fork body to detect capacitance changes associated with ingestion. (b) Illustration of a diner using InBite across three food-ingestion phases to sonify the evolving sounds.

## Abstract

Sound can enhance taste perception and the overall dining experience. Accordingly, HCI researchers have begun designing auditory dining systems that connect diners with food to support more immersive multisensory experiences. However, prior work has largely overlooked how continuous auditory contexts (e.g., genre, timbre, and pitch) can resonate with the bite-by-bite multisensory dynamics

of eating. Thus, we present InBite, an interactive fork and companion app developed through three iterative design cycles. InBite enables diners to configure their auditory experience in real time during meals: via the app, diners can generate AI music grounded in their momentary sensory experience and transfer it to the fork for playback. The fork then senses ingestion trajectories and intake states to dynamically modulate the sound in response to the sensory experience of each bite. Through this work, we contribute design knowledge for bite-level modulation of auditory dining experiences in Human-Food Interaction research.



This work is licensed under a Creative Commons Attribution 4.0 International License.  
*CHI EA '26, Barcelona, Spain*  
© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2281-3/26/04  
<https://doi.org/10.1145/3772363.3799269>

## CCS Concepts

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

## Keywords

Human-food interaction, auditory interface, sound

### ACM Reference Format:

Hongyue Wang, Yuting Wu, Subasinghe Piyaathnage Sahan Madusanka, Ziqi Fang, Yuchen Zheng, Hong Luo, Jialin Deng, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2026. InBite: Towards Designing Real-time Configurable Auditory Dining Experience for Each Bite. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3772363.3799269>

## 1 Motivation and Background

Human–Food Interaction (HFI) [3–6, 13, 25, 29–31], an emerging subfield of Human–Computer Interaction (HCI), has continually explored how digital technologies can enrich dining experiences, with particular attention to how digital stimuli can be synchronized with the act of eating to reshape the sensory aspects of food. Recent research has begun to design bite-level modulation of eating experiences, which refers to adjusting diners' perception through variable digital stimuli across different phases of food ingestion (e.g., taking a bite, chewing, and swallowing). Existing work, on the one hand, modulate specific taste dimensions, such as saltiness or sourness, through intraoral interventions [14, 19] (e.g., electrical stimulation of taste); on the other hand, through extraoral cues (e.g., olfactory release [9] and visual augmentation [8, 18]) that layer additional sensory information onto the eating process.

Nevertheless, current implementations of bite-level experiences in HFI are largely limited to a discrete “bite-and-respond” paradigm: a bite is treated as a single trigger, upon which a pre-defined stimulus is presented. This approach, to some extent, overlooks that dining unfolds as a dynamic, time-extended process in which each bite is jointly shaped by multisensory food cues (e.g., smell, visual appearance, and texture), eating intentions, and shifts in eating rhythm [17, 28]. Living Bento [2], for example, uses diners' heart-rate variations to drive dynamic lighting effects on Kudzu Jelly Noodles, highlighting that dining states can evolve over the course of a meal and be continuously modulated through digital outputs. However, whether relying on passive mappings from physiological signals or on pre-scripted digital stimuli, prior work has generally lacked an active modulation space that enables diners to form a stable association between their eating behavior and the system's output. Taken together, this suggests that meaningful bite-level sensory modulation rely not only on the timing and modality of added stimuli, but also on how a system maps diners' intentions and evolving experiential states onto digital outputs that are intelligible and actively engageable.

Building on these challenges, we turn to sound as a ubiquitous sensory element in dining. Crossmodal research has shown that digital sound during eating can enhance the perception of taste and texture [7], thereby reshaping diners' food experiences [11]. However, in existing sound-related HFI work, auditory outputs are typically delivered either as background music or as pre-programmed audio clips that are triggered when eating actions (e.g., food manipulation [23] or licking [22, 24]) are detected. We argue that the auditory space of sound (e.g., pitch, rhythm, and density) remains

underexplored as a means of enriching momentary dining experiences. Moreover, these approaches offer limited support for diners to actively modulate the auditory experience in response to their evolving sensations while eating.

With advances in human–AI sound co-creation, contemporary generative models can condition on various inputs (e.g., modalities, parameters, and weighting configurations) to produce creative music outputs that reflect users' intentions. This process parallels how diners perceive and consume food, where each bite presents a unique combination of ingredients and multisensory interactions that shape a dynamic dining experience. This parallel inspired us to explore how diners might meaningfully reconfigure sound during eating to align auditory information with their distinctive, in-the-moment bite-level experience. Meanwhile, to facilitate diners' agency in actively shaping their bite-level auditory dining experience, we adopted sonification [10], which maps data derived from diners' food-ingestion behaviors onto perceptible changes in acoustic parameters through real-time modulation<sup>1</sup>. Motivated by these insights, we propose the research question below: *How can we design a real-time configurable auditory dining experience that responds to the dynamic changes of multisensory experience for each bite?*

To address this question, we present InBite, an AI-powered sonification fork and companion app that enables diners to generate music aligned with their momentary sensory changes during eating via voice input, and to listen to the resulting output directly through the fork. InBite senses food-ingestion trajectories and intake states, and maps these signals to bite-level sonification of the melody. Overall, this work contributes design knowledge—derived from three iterative design rounds—on how to modulate auditory dining experiences for each bite, which can inspire the HFI community to leverage digital elements to design bite-level modulation that aligns with diners' moment-to-moment multisensory perceptions during eating.

## 2 System Design

We adopt Research through Design (RtD), a reflective practice that generates new knowledge and understanding through the conception, design, iteration, and experimentation of novel artifacts [32]. In this section, we present our reflections from a two-version iterative process of designing InBite.

### 2.1 Exploring real-time configurable auditory dining experience for each bite

*2.1.1 InBite-v1: Design and Reflection.* Prior HFI research offers limited design knowledge for dynamically shaping bite-level dining experiences. Thus, we aim to explore such design to align auditory information with the momentary multisensory shifts that occur during eating, enabling more emotionally resonant digital dining experiences[33]. Toward this goal, we developed the InBite-v1 [21] interaction in three aspects:

<sup>1</sup> Sonification is a design approach that transforms data into perceptible relations within an acoustic signal, offers a principled way to construct such a continuous control space through parameter modulation.

**System Design:** We developed InBite-v1 (Figure 2) using an ESP32-S3 development board. We connected the ESP32-S3’s capacitive sensing circuit to the stainless-steel fork body and housed the electronics in a custom 3D-printed enclosure, enabling the system to capture changes in capacitance during food ingestion. We selected capacitive sensing because it can reflect key ingestion states (e.g., contact with the lips and the inside of the mouth), indicating when food is being eaten and providing actionable inputs for bite-level sonification. Finally, we integrated a speaker at the end of the fork, as prior work suggests that emitting sound from the interactive artifact itself can strengthen diners’ affective connection with the food [23, 24].



**Figure 2: Prototyping process of InBite-v1:** (a) feasibility testing on a breadboard; (b) final prototype on a stainless-steel fork; (c) assembly within a 3D-printed enclosure; (d) a person uses InBite-v1 to eat a piece of apple, triggering sounds. (e) the human-AI sound co-creation workflow in ComfyUI. (f) App interface for receiving diners’ text input to generate AI music that represents their momentary dining perceptions.

**Sound Creation:** We implemented a human–AI sound co-creation workflow in ComfyUI using the Stable Audio model and deployed it in a mobile app that enables music generation via text prompts. We also provide an optional integrated AI agent (Appendix A). Beyond recognizing diners’ momentary intentions [26, 27], the agent refines diners’ initial descriptions of in-the-moment sensory experiences into more effective prompts (e.g., genre, BPM, and instrument types), thereby increasing prompt specificity and improving the generative model’s output quality. To minimize semantic interference during eating, we primarily used instrumental music. This co-creative pipeline allows diners to iteratively adjust the generated music in response to each bite’s multisensory experience. The selected output is then transferred to InBite-v1 via Wi-Fi for playback when the fork touches the diner’s lips. We sensed changes in capacitance and mapped the resulting variation range to dynamic volume modulation of the playback audio.

Through the design process, we identified two challenges in constructing real-time configurable auditory experience for bite-level modulation: (1) **Limited sonification parameters.** On the ESP32-S3, we only modulated audio volume per bite, which often felt monotonous. Future iterations should expand sonification dimensions and mappings to increase expressive variation for each bite. (2) **Missing onset–offset anchors for sonification.** The food-handling process prior to ingestion can provide intentional cues that anticipate an upcoming bite. For example, forcefully cutting a steak may indicate heightened appetite and anticipation. Future iterations should therefore move beyond sensing discrete intake states and instead capture the full ingestion trajectory, from bite onset (e.g., pickup untential and transport food into mouth) to offset

(e.g., put utensil back), to support a more complete representation of the dining process.

**2.1.2 InBite-v2: Design and Reflection.** Based on the above, we designed a second iteration.

**System Design:** First, we extended our sonification mappings in Max/MSP beyond volume control to include additional parameters—velocity, pitch, and tempo—which are widely used in prior HCI work [15, 20]. To stream audio via A2DP, we used a Raspberry Pi Zero 2 and integrated an inertial measurement unit (IMU), a capacitive sensor, and a speaker module. The IMU and capacitive sensor were used to define, respectively, the onset and offset of food-ingestion trajectories and the intake states used for sonification. To mount the Raspberry Pi at the end of the fork, we added (1) a rechargeable power module to eliminate tethering cables and (2) a boot script that automatically starts sensor acquisition on power-up and transmits data to Max/MSP over Wi-Fi for real-time sonification.



**Figure 3: Prototyping process of InBite-v2:** (a,b) The front and back of the sensing module; (c) final prototype on a stainless-steel fork; (d) The Max/MSP interface for sonification experience design.

**Sound Creation:** Diners can describe their momentary sensory experience to prompt AI-generated music via app. We revised the workflow to produce two corresponding melodic layers with different instruments, *Melody A* mapped to IMU dynamics and *Melody B* to capacitance changes, while keeping the seed<sup>2</sup> parameter constant across generations. This design encourages the two layers to share a coherent musical structure and remain harmonically compatible when played simultaneously. In Max/MSP, we further implemented seamless music updating: each time a diner prompts new music generation, the system smoothly replaces the current audio file to support a real-time, configurable auditory dining experience.

**Sonification:** We first ran a 10s calibration to record baseline values for acceleration, angular velocity, and capacitance. We then detected whether the participant picked up InBite by checking whether instantaneous tri-axial acceleration deviated from baseline by more than 20%. During food processing, we used the acceleration magnitude  $E(t)$ <sup>3</sup> to represent movement intensity and mapped it to modulate *Melody A*. Similarly, when capacitance exceeded its baseline by more than 20% (indicating food intake), we mapped capacitance changes to modulate *Melody B*. Finally, when data returned to baseline and remained stable within 5s, we treated this as the end of one food ingestion.

<sup>2</sup>A random seed is an initialization value that controls stochastic generation, enabling the same prompt and settings to reproduce identical or highly similar AI-generated outputs.

<sup>3</sup> $E(t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2}$ , where  $a_x(t)$  denotes the acceleration along the  $x$ -axis at time  $t$  (similarly for  $a_y(t)$  and  $a_z(t)$ ).

Throughout the design process, we identified two artifact-level challenges: (1) **Balancing functionality and form factor**. Adding sensors can expand sonification, but also increases size and weight, potentially disrupting eating movements. The design must therefore maximize utility while minimizing bulk. (2) **Limited grip comfort**. The current enclosure prioritizes component housing over natural grasping and continuous movement. Its form should better support tool embodiment [12], keeping attention on the food rather than device operation.

**2.1.3 Design considerations from RtD process.** Building on the design practice above, we derived three design considerations (DCs) on system functionality, sonification mappings, and interaction form.

**DC1: Design continuous sonification through live audio-stream modulation.** Inspired by Singh et al.'s Go-with-the-Flow framework [16], which highlights how three movement anchors (initial posture, a movement trajectory, and a final posture) can be mapped to sonification elements, we argue that bite-level sonification should be experienced as a continuous process rather than a discrete trigger. Accordingly, InBite senses key food-ingestion anchors (pickup, trajectory, and intake) to modulate a live audio stream and deliver a dynamic sound through the artifact.

**DC2: Design perceivable sonification mappings that support diners' agency in evolving the sound.** Sonification mappings should help diners develop a clear sense of agency [1]: their movements should produce stable, recognizable changes in sound so that they feel they are actively shaping the sound rather than passively receiving feedback. Accordingly, the priority is not to add more parameters, but to make the most consequential differences perceptible.

**DC3: Design non-disruptive interactions that preserve the natural flow of dining.** Interactions should minimally interrupt the natural eating rhythm. First, sensing modules should align with everyday cutlery conventions in size, weight, and center-of-mass distribution, and be integrated into InBite to avoid conspicuous hardware or intrusive cables. Second, audio should originate from InBite itself rather than external devices, strengthening perceived resonance between eating and the auditory experience.

## 2.2 Final Prototype

**2.2.1 Hardware design.** To enable more natural everyday dining interactions (DC3), we consolidated InBite's electronics into a custom PCB centered on an ESP32-WROOM-32, which supports Bluetooth A2DP audio streaming, Wi-Fi/BLE, and capacitive sensing (DC1). This reduced size, improved robustness, and fit sensing, computation, wireless communication, and audio output within the fork handle. The PCB integrates an MPU6050 IMU, a MAX98357A I2S class-D amplifier, and a miniature speaker, plus an on/off switch and power-management circuitry for portable battery operation. We designed an ergonomically contoured handle in SolidWorks and 3D-printed it in food-safe PETG to house the components while aiming to maintain aesthetics (DC3).

**2.2.2 AI music generation.** We updated the companion app to support voice prompts for AI music generation. We also structured

the generation into three tracks: a base main melody and two accompaniment layers (Chord A and Chord B) derived from the main track using the same seed. Chord A and Chord B respond to food-ingestion trajectories and states, respectively. All tracks were generated as solo-instrument melodies to make each interaction's impact perceptible.

**2.2.3 Sonification.** InBite and the host server running Max/MSP were connected to the same local network. Sensor streams (acceleration:  $acc_x$ ,  $acc_y$ ,  $acc_z$ ; angular velocity:  $ang_x$ ,  $ang_y$ ,  $ang_z$ ; capacitance:  $cap$ ) were transmitted to Max/MSP for sonification mapping and modulation. The rendered audio stream was then routed back to InBite for playback via Bluetooth A2DP. To make sonification differences perceptible (DC2), we used reverb as a spatial parameter in addition to tempo, pitch, and velocity. To mitigate auditory discomfort in dining contexts [20], we excluded very low (20–200 Hz) and high (>1,000 Hz) ranges from the primary pitch-modulation region. Our sonification algorithm comprises three stages. Each stage specifies a trigger condition, target musical layer, and mapping inputs, enabling feedback that is both continuous and structurally organized. The workflow is summarized below:

**Stage 1: Calibration.** (1) *Preprocessing.* Upon receiving raw sensor data from InBite, the host performs coordinate alignment, EMA smoothing to suppress high-frequency noise, and outlier removal to reduce transient spikes. (2) *Static calibration.* At the beginning of the meal, diners place InBite on the table and switch it on. The system then runs a 10 s calibration window and computes mean baselines ( $acc_0$ ,  $ang_0$ ,  $cap_0$ ). After calibration, the main melody starts as a continuous background layer. (3) *Pickup detection.* We detect pickup by monitoring the smoothed  $a_z$  and its deviation from the baseline. If the deviation exceeds 20% for at least 200 ms, the fork is classified as picked up and Chord A is triggered with a 1 s fade-in to avoid abrupt onset.

**Stage 2: Food-ingestion trajectory sonification.** After pickup, the system continuously overlays Chord A on the main melody and computes movement intensity as the acceleration magnitude  $E(t)$ . We smooth  $E(t)$  (e.g., EMA) to suppress tremor-induced jitter, then normalize it to  $[0, 127]$  to modulate tempo, pitch, velocity, and reverb depth.

**Stage 3: Food-ingestion sonification.** (1) *Bite onset.* When food reaches the mouth, we overlay Chord B to create an auditory climax. This stage is triggered by capacitance: if  $cap(t)$  deviates from baseline by more than an ingestion threshold  $\theta_{ingest}$ , we detect bite onset and map  $cap(t)$  to normalized  $[0, 127]$  parameters for real-time modulation. (2) *Sustained ingestion.* During ingestion, all three layers (main melody + Chord A + Chord B) continue to play, while ongoing changes in  $cap(t)$  further increase tempo, pitch, velocity, and reverb depth. (3) *Termination.* When InBite is placed on the table and all data velocity return to baseline and remain stable for 5 s, the episode ends. Chord A and Chord B fade out over 1 s, while the main melody remains. Users can stop all output by switching off the device.

### 3 Discussion, Limitations and Conclusion

In recent digital dining experiences, DJs have modulated auditory dining experience in real time based on the rhythm of on-site eating.<sup>4</sup> While such performative orchestration supports bite-level variation across a meal, it is driven by external experts and often abstracts away what diners eat, how they perceive food moment to moment, and how situated eating actions unfold. Yet these situated, bite-level dynamics are critical for grounding sound in food perception. Across InBite's iterative design, we aimed to provide a configurable, continuously responsive sound mechanism that lets diners listen, perform, and steer the musical flow through everyday eating. We pursue two complementary directions. First, music generation should reflect diners' in-the-moment multisensory experience. Thus, we explore human-AI co-creation to couple musical content with situated food perception, building on prior HFI work on synchronizing sound and eating behavior [23]. We extend this coupling beyond discrete trigger-response toward continuous creation, switching, and modulation of musical parameters (e.g., genre, timbre, tempo) in response to bite-level sensory changes. Second, sound performance should be driven by each diner's behavior, from ingestion trajectories to intake states, so diners actively shape the evolving sound rather than consume pre-programmed audio. Our work has several limitations. First, diners currently use a mobile app to provide voice prompts; future iterations could integrate a miniature microphone into the PCB to reduce interaction friction. Second, we have not yet evaluated InBite in situ; a restaurant deployment could examine real-world use and adoption.

### Acknowledgments

We sincerely thank all the reviewers for their feedback and contributions. Florian 'Floyd' Mueller thanks the Australian Research Council, especially DP190102068, DP200102612, and LP210200656.

### References

- [1] Joanna Bergström, Jarrod Knibbe, Henning Pohl, and Kasper Hornbæk. 2022. Sense of agency and user experience: Is there a link? *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–22.
- [2] 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*. 1–18.
- [3] 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
- [4] Jialin Deng, Patrick Olivier, and Florian 'Floyd' Mueller. 2021. Design of cyber food: beginning to understand food as computational artifact. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–6. doi:10.1145/3411763.3451687
- [5] Ziqi Fang, Hongyue Wang, Nilakna Disiawari Warushavithana, Subasinghe Pi-yarathnage Sahan Madusanka, Jialin Deng, Weijen Chen, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2026. The Wizard of Taste: Demonstrating Multisensory Eating with Electrical Taste and Bone-conduction Auditory Seasoning. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. 1–5. doi:10.1145/3772363.3799379
- [6] Florian 'Floyd' Mueller, Tim Dwyer, Sarah Goodwin, Kim Marriott, Jialin Deng, Han D. Phan, Jionghao Lin, Kun-Ting Chen, Yan Wang, and Rohit Ashok Khot. 2021. Data as delight: Eating data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3411764.3445218
- [7] David Guedes, Margarida Vaz Garrido, Elsa Lamy, Bernardo Pereira Cavalheiro, and Marília Prada. 2023. Crossmodal interactions between audition and taste: A systematic review and narrative synthesis. *Food Quality and Preference* 107 (2023), 104856. doi:10.1016/j.foodqual.2023.104856
- [8] Saumya Gupta, Ferran Altarrriba Bertran, Ödüz 'Oz'Buruk, Sara Milkes Espinosa, Theresa Jean Tanenbaum, and Minerva Wu. 2021. Exploring Food based Interactive, Multi-Sensory, and Tangible Storytelling Experiences. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 651–665. doi:10.1145/3461778.3462006
- [9] Osama Halabi and Mohammad Saleh. 2021. Augmented reality flavor: cross-modal mapping across gustation, olfaction, and vision. *Multimedia Tools and Applications* 80, 30 (2021), 36423–36441.
- [10] Thomas Hermann. 2008. Taxonomy and definitions for sonification and auditory display. In *Proceedings of the 14th international conference on auditory display*, Vol. 8. Center of Excellence-Cognitive Interaction Technology CITEC Bielefeld.
- [11] Rebecca Kleinberger, Akito Oshiro Van Troyer, and Qian Janice Wang. 2023. Auditory seasoning filters: Altering food perception via augmented sonic feedback of chewing sounds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [12] Marie Martel, Lucilla Cardinali, Alice C Roy, and Alessandro Farnè. 2016. Tool-use: An open window into body representation and its plasticity. *Cognitive neuropsychology* 33, 1-2 (2016), 82–101.
- [13] Marianna Obrist, Ferran Altarrriba Bertran, Neharika Makam, Soh Kim, Christopher Dawes, Patrizia Marti, Maurizio Mancini, Eleonora Ceccaldi, Nandini Pasmurthy, Sahej Claire, et al. 2024. Grand challenges in human-food interaction. *International Journal of Human-Computer Studies* 183 (2024), 103197. doi:10.1016/j.ijhcs.2023.103197
- [14] Nimesha Ranasinghe, David Tolley, Thi Ngoc Tram Nguyen, Liangkun Yan, Barry Chew, and Ellen Yi-Luen Do. 2019. Augmented flavours: Modulation of flavour experiences through electric taste augmentation. *Food Research International* 117 (2019), 60–68.
- [15] Richard Savery, Madhukesh Ayyagari, Keenan May, and Bruce Walker. 2019. Soccer sonification: Enhancing viewer experience. In *International Conference on Auditory Display (25th: 2019)*. Georgia Tech, 207–213.
- [16] Aneeha Singh, Stefano Piana, Davide Pollarolo, Gualtiero Volpe, Giovanna Varni, Ana Tajadura-Jiménez, Amanda CdeC Williams, Antonio Camurri, and Nadia Bianchi-Berthouze. 2016. Go-with-the-flow: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. *Human-Computer Interaction* 31, 3-4 (2016), 335–383.
- [17] Dale Southerton, Cecilla Diaz-Méndez, and Alan Warde. 2012. Behavioural change and the temporal ordering of eating practices: A UK–Spain comparison. *International Journal of Sociology of Agriculture and Food* 19, 1 (2012), 19–36.
- [18] Charles Spence. 2013. Multisensory flavour perception. *Current Biology* 23, 9 (2013), R365–R369. doi:10.1016/j.cub.2013.01.028
- [19] Asif Ullah, Yifan Liu, You Wang, Han Gao, Hengyang Wang, Jin Zhang, and Guang Li. 2022. E-Taste: Taste sensations and flavors based on tongue's electrical and thermal stimulation. *Sensors* 22, 13 (2022), 4976.
- [20] Concepción Valdez, Katya Alvarez-Molina, Luis A Castro, and Monica Tentori. 2023. Zens: Designing and evaluating a movement sonification interface to alter body and flexibility perception in Yoga. *International Journal of Human-Computer Studies* 178 (2023), 103084.
- [21] Hongyue Wang, Sasindu Abewickrema, Hong Luo, Yuchen Zheng, Shan Luo, Ziqi Fang, Jialin Deng, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2025. SoniFork: Towards Chef-AI Sound Co-Creation for Orchestrating Real-Time Auditory Dining Experiences. In *Proceedings of the 2025 International Conference on Human-Engaged Computing (ICHEC 2025)*. 1–6. doi:10.1145/3786995.3787073
- [22] Hongyue Wang, Sasindu Abewickrema, Yuchen Zheng, 'Cosmos' Po-Yao Wang, Hong Luo, Ziqi Fang, Jialin Deng, Nandini Pasmurthy, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2026. GastroConcerto: Towards Designing Dining-Sound Pairings to Support Culinary Creativity. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. 1–23. doi:10.1145/3772318.3790348
- [23] Hongyue Wang, Jialin Deng, Linjia He, Nathalie Overdeest, 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. *Proceedings of the 2025 CHI conference on human factors in computing systems (2025)*, 1–19. doi:10.1145/3706598.3714237
- [24] Hongyue Wang, Jialin Deng, Dehui Kong, Ziqi Fang, Hong Luo, Nandini Pasmurthy, Rakesh Patibanda, Sasindu Abewickrema, Xiao Zoe Fang, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2025. GastroConcerto: Towards Designing Auditory Dining System to Enrich Chefs' Culinary Practices. 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 298, 8 pages. doi:10.1145/3706599.3719700
- [25] Hongyue Wang, Jialin Deng, Aravind Mohan, Yinyi Li, Hao Peng, Linjia He, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2024. pic2eat: Facilitating Social Ice-breaking through Collaborative Design of 3D Printed Appetizer. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7. doi:10.1145/3613905.3651082

<sup>4</sup><https://service95.com/food-music-experience-paired>

- [26] Hongyue Wang, Zhiqian Feng, Jinglan Tian, and Xue Fan. 2022. MFA: a smart glove with multimodal intent sensing capability. *Computational Intelligence and Neuroscience* 2022, 1 (2022), 3545850. doi:10.1155/2022/3545850
- [27] Hongyue Wang, Zhiqian Feng, Xiaohui Yang, Liran Zhou, Jinglan Tian, and Qingbei Guo. 2024. MRLab: Virtual-reality fusion smart laboratory based on multimodal fusion. *International Journal of Human-Computer Interaction* 40, 8 (2024), 1975–1988. doi:10.1080/10447318.2023.2227823
- [28] Alan G Watts, Scott E Kanoski, Graciela Sanchez-Watts, and Wolfgang Langhans. 2022. The physiological control of eating: signals, neurons, and networks. *Physiological reviews* 102, 2 (2022), 689–813.
- [29] 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
- [30] Yuchen Zheng, Ashlyn Doyle, Hongyue Wang, Jialin Deng, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2025. "When You Are the Plant": Encouraging Hydration Through Playful Virtual Reality Embodiment. In *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 793–794. doi:10.1109/ISMAR-Adjunct68609.2025.00193
- [31] 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)*. 1–23. doi:10.1145/3772318.3790615
- [32] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502. doi:10.1145/1240624.1240704
- [33] Amit Zoran. 2019. Cooking with computers: The vision of digital gastronomy [point of view]. *Proc. IEEE* 107, 8 (2019), 1467–1473.

Spacey Synths, Moog Oscillations, TR-909 Drum Machine, 808 Hip Hop Beat, Funk Drums, Tabla, Steel Drum, Koto, Pipa, Sitar, Cello, Woodwinds, Alto Saxophone, Trumpet, Vibraphone, Glockenspiel, Drumline, Djembe Genres (examples): Lo-Fi Hip Hop, Ambient, Orchestral Score, Jazz Fusion, Bossa Nova, Deep House, Trip Hop, Indie Folk, Indie Electronic, Afrobeat, Indian Classical, Synthpop, Techno, Drum & Bass, Vaporwave, Celtic Folk, Latin Jazz, Reggae Mood/Style tokens (examples): Warm, Nostalgic, Playful, Mysterious, Refined, Intimate, Energetic, Dreamy, Ominous, Bright Tones, Tight Groove, Ethereal Ambience, Subdued Melody, Crunchy Distortion, Danceable, Experimental, Rich Orchestration

Final Check Before Output One line only. Exact format and separators. 3–5 words per field.

## A Appendix: Description of the Prompting Framework

You are an AI assistant that converts a diner's dish description into a single, concise prompt optimized for Stable Audio. Your output must follow the exact structured format below and contain no extra text.

Role Input: A diner or chef provides an abstract or concrete description of a dish, ingredients, culture, mood, texture, temperature, and/or feeling. Task: Translate that description into a simple, structured music prompt that matches the intended dining experience.

Output Format [Band/Solo] | Genre: [Main Genre] | Subgenre: [Optional] | Instruments: [Key Instruments] | Moods: [Emotions] | BPM: [Tempo] | Style: [Ambiance]

Strict Output Rules Return exactly ONE line in the specified format. Do not add explanations, labels, bullet points, examples, or multiple options. Each field must be 3–5 words maximum. Subgenre is optional; if not needed, omit the entire "| Subgenre: ..." segment. Do not output an Atmosphere field or any additional fields.

Interpretation Guidelines Identify the diner's intent (comforting, edgy, fresh, indulgent, nostalgic, refined, chaotic, etc.). Map ingredients/culture to an appropriate genre and instrument palette. Translate texture/temperature to sound qualities: crisp → tight groove, bright tones; creamy → smooth pads, warm melody; spicy → driving rhythm, sharper timbres; smoky → darker tones, saturated textures; citrus/fresh → bright, airy, light percussion. If the input is vague, enrich minimally by selecting fitting genre, a small set of instruments, and a coherent mood. Choose BPM to match pacing: slow (60–90) for calm/intimate; mid (90–120) for conversational/dining flow; fast (120–150) for energetic/street-food/party.

Allowed Descriptor Bank (optional use) Instruments (examples): Warm Acoustic Guitar, Smooth Pianos, Rhodes Piano, Synth Pads,