SoniFork: Towards Chef-AI Sound Co-Creation for Orchestrating Real-Time Auditory Dining Experiences

Hongyue Wang*
hongyue@exertiongameslab.org
Exertion Games Lab, Monash University
Melbourne, Victoria, Australia

Sasindu Abewickrema*
sasindu.jananga@gmail.com
Exertion Games Lab, Monash University
Melbourne, Australia

Hong Luo
hong.luo@monash.edu
Exertion Games Lab, Monash University
Australia

Yuchen Zheng
Exertion Games Lab, Monash University
Melbourne, Victoria, Australia

Shan Luo
luoshan3398@qq.com
China Academy of Art
Hangzhou, Zhejiang, China

Ziqi Fang
ziqi@exertiongameslab.org
Exertion Games Lab, Monash University
Australia

Jialin Deng
jialin.deng@bristol.ac.uk
School of Computer Science, University of Bristol
United Kingdom

Don Samitha Elvitigala
don.elvitigala@monash.edu
Exertion Games Lab, Monash University
Melbourne, Victoria, Australia

Florian 'Floyd' Mueller
floyd@exertiongameslab.org
Exertion Games Lab, Monash University
Melbourne, Victoria, Australia

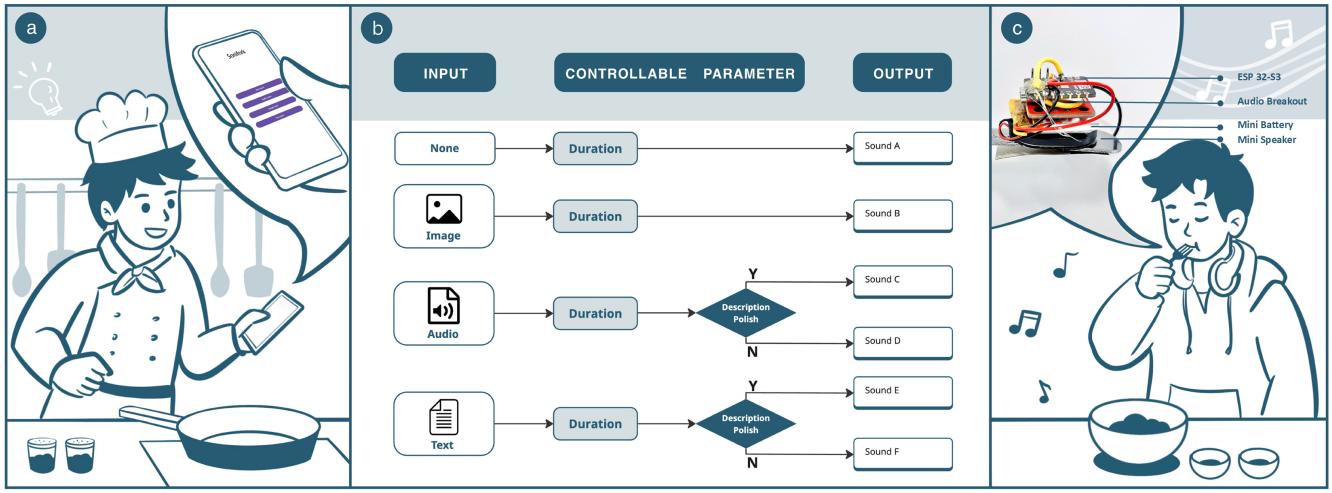

Figure 1: Usage Flowchart of Sonifork. (a) A chef uses our app to generate AI sounds. (b) Four chef-AI sound co-creation workflows to generate AI sound. (c) A diner is using SoniFork to enjoy an auditory dining experience with each bite.

Abstract

Sound plays a crucial role in shaping eating experiences and sensory perception. This has drawn increasing attention from the HCI community to use sound as a means to enrich chefs' culinary creativity. However, ownership over sound creation has remained with interaction designers or restaurateurs, overlooking how sound can represent chefs' authentic culinary expressions. With the advancements of human–AI sound co-creation, contemporary AI models allow users to modify various inputs (e.g., modality, parameters, and weights) to generate creative outputs. This process closely mirrors how chefs experiment with ingredients (e.g., combinations, proportions, and layers) to craft curated dining experiences. This parallel inspired us to explore the the opportunity for chefs to meaningfully reconfigure the sound within their dishes, aligning the auditory elements with their distinctive culinary vision. To demonstrate this, we created SoniFork, a fork-like system with capacitive sensing and sound-release capabilities, that lets chefs orchestrate real-time auditory experience via its companion app at each bite. This work

*Both authors contributed equally to this research.

ICHEC 2025, Singapore, Singapore

ACM ISBN 979-8-4007-2234-9/2025/11
https://doi.org/10.1145/3786995.3787073

repositions sound design ownership within the kitchen, supporting chefs' culinary creativity through chef-AI sound co-creation.

CCS Concepts

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

Keywords

Human-food interaction, auditory interface, culinary creativity

ACM Reference Format:

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), November 21–23, 2025, Singapore, Singapore.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3786995.3787073

1 Introduction

In the evolving catering industry, culinary practitioners curate not only the flavors of their dishes but also the entire dining experience. Their creative expression, manifested through plating, temperature, aroma, and meal pacing, deeply influences how people engage with food. Yet, the potential of sound as a novel culinary resource within this creative practice remains underexplored. Traditionally, sound has been shown to play a crucial role in shaping eating experiences [34], altering sensory perception [35], and influencing eating behaviors [37]. These findings have sparked growing interest within the field of Human-Food Interaction (HFI) [8, 9, 13, 29, 33, 42, 43, 51], where sound is increasingly being explored as a means of enhancing culinary creativity [41, 42].

However, the production and arrangement of sound are typically handled by stakeholders such as interaction designers [48] and restaurateurs [4]. These stakeholders often need to invest considerable time in designing creative sound concepts and inventories for practical deployment. This reliance on predefined auditory sources constraints chefs in their creative process [26], reinforcing the stereotype that they cook for themselves rather than for diners [50]. This observation highlights an opportunity to shift the ownership of sound creation from these stakeholders to chefs, enabling a more authentic representation of their culinary vision. Especially, the advancements in AI capabilities are gradually narrowing the knowledge gap between audio experts and novices (e.g., Chefs) [28, 31], highlighting the potential of human-AI sound co-creation to further enhance creative exploration. Thus, this paper makes the following contribution: (1) an app integrated four chef–AI sound co-creation workflows, enabling chefs to create personalized sounds that represent their culinary vision within a dish; (2) *SoniFork* system was developed to enable chefs to transmit and adjust AI-generated sounds in real time and transform diners' unique eating behaviors into dynamic auditory experiences. By sonifying diners' embodied interactions, the system supports chefs in crafting evolving soundscapes that reflect their culinary intent in each bite and enhance the sensory depth of the dining experience.

2 Related Work

2.1 Auditory Interaction in Human-Food Interaction

Sound elements in dining leverages crossmodal effects to enhance or alter the perception of taste and texture [17], thereby reshaping diners' food experiences. Intrinstic sounds during eating (e.g., the amplified, high-frequency "crunch" of a potato chip) can significantly increase subjective evaluations of crispness and freshness [39]. Meanwhile, environmental sounds can modulate perceptions of flavour and mouthfeel: high-pitched music tends to enhance sweetness, whereas low-pitched tones are more likely to accentuate bitterness [36]. Inspired by these findings, the Human–Food Interaction (HFI) community has progressively repositioned sound from a "background soundtrack" to a responsive interaction design element. For instance, 'iScream' employs capacitive sensing to produce incongruent sonic responses to licking gestures, thereby enhancing playfulness and engagement [49]. Moreover, several studies showcased that auditory interactions can support mindful eating or reduce snack consumption [30]. Weijen et al. [6] used audible frequencies to drive liquid-surface patterns mapped to taste categories, thereby enhancing taste perception through visual and tactile cues and influencing texture perception, appetite, and enjoyment. However, the dynamic role of sound within culinary practice in shaping dining experiences remains underexplored. For example, augmented-reality cooking assistants [40] combine visual and auditory notifications (e.g., simulated bubbling) to guide cooking, yet their emphasis largely remains on task completion and procedural correctness rather than on chefs' expression in orchestrating diners' sonic experiences. Then, Wang et al. [41, 42] collaborated with chefs to explore links between food creation and sonic expression; while sound enhanced their culinary creativity, limited musical expertise imposed a substantial music-design burden on chefs during creation.

To bridge this gap, we present a chef–AI music co-creation workflow enabling chefs to orchestrate the soundscape of each bite according to their authentic culinary intent.

2.2 Empowering Computational Creativity in Culinary Practices via Human-AI Sound Co-creation

Machine learning models have long been applied to music creation [12, 44, 47, 52], from early approaches in symbolic music composition (e.g., MIDI) [19, 21] to more recent advancements in audio generation [1, 7], continually driving the evolution of co-creative music systems. Some systems are designed to collaborate with professional musicians [23], offering precise control that aligns with their creative goals. These systems often emphasize AI responsiveness to user input within well-specified creative constraints, ensuring coherence between system behavior and the user's intent [2, 23, 25, 38]. In contrast, other systems respond to users' contextual cues or ambient behaviors, generating content that dynamically adapts to the user's evolving creative process [14, 15, 24]. These sound co-creation approaches reflect a dynamic interplay of agency between AI-driven and user-directed forms of creativity. This distinction is also evident in the culinary practices, where creative

expression stems not only from standardized requirements (e.g., taste or ingredient preparation) but also from aesthetic improvisation, such as plating and culinary performance. The complexity of chefs' interactive intentions and needs invites a reconsideration of how to meaningfully integrate AI's computational creativity within the culinary domain. Building on prior interaction modes from human-AI sound co-creation, we propose that future chef-AI sound co-creation should focus on the following four aspects:

AI autonomy in creativity. With the emergence of large-scale generative models such as MusicLM [1] and Jukebox [10], AI-driven sound generation has become increasingly autonomous, wherein music emerges not as a direct response to human instruction, but as a result of the system's own generative structure. Researchers have identified this model-led unpredictability as a source of creativity [5]. In this context, chefs may fully delegate the auditory component to the AI, aligning their culinary creation with the music produced by the system. This approach is particularly well-suited for chefs who lack a concrete auditory vision but are eager to integrate sound into their gastronomic expression.

Context-responsive co-creation. Recent advancements in AI have positioned it as a perceptual hub capable of interpreting non-verbal signals for real-time and improvised sound generation. Models such as EMOPIA [22] enable emotion-driven music synthesis, while various HCI studies have explored audio models that dynamically respond to visual input [14], gestural cues [24], and real-time dance choreography [15]. Thus, chefs can incorporate their culinary context, such as the process or creative outcome, as input for sound co-creation, facilitating a deeper sensory integration between food creation and sound.

Audio-informed co-creation. Audio inputs inherently carry rich musical information (e.g, genre, loudness, and tempo) enabling a range of novel sound generative techniques, including tone transfer [11], genre transformation [3], and the creation of harmonies from existing audio [20]. As some chefs might demonstrate more open-ended sonic intentions for their culinary presentations, this approach allows them to generate soundscapes from musical pieces or even from their own musical input (e.g., vocal humming), embedding both their personal aesthetic and existing auditory art forms into the overall dining experience.

Chef autonomy in creativity. Unlike exploratory or improvisational co-creation, text-to-audio generation models such as Suno [2] or Udio [38] enable users with musical sensibilities to craft precise auditory outputs (e.g., music, sound effects) through explicit textual prompts. This approach is particularly suitable for chefs who assume full authorship over the multisensory experience, such as in omakase-style dining, where chefs determine both what and how diners eat. In this context, the role of generative AI in the co-creative process functions primarily as a responsive tool that extends the chef's creative intent, thereby facilitating a clear conceptual mapping between sound creation and dish design.

In summary, our review of prior work reveals a knowledge gap in sound-empowered culinary practices that could be addressed through chef–AI sound co-creation. To explore this direction, this paper proposes four co-creative processes (i.e., no input, visual input, audio input, and textual input), each representing a distinct creative role that empowers chefs to take ownership in orchestrating sonic dining experiences.

3 System Design

3.1 Design Considerations

We aim to map the interaction pathways for chef-involved auditory interaction through practice-based, reflective design process [16, 18, 27, 45, 46]. Therefore, we investigate two contrasting AI-based music-generation paradigms (Figure 2): (1) **Live, interactive neural-audio processing:** We deploy the RAVE model in Pure Data to perform low-latency, continuously controllable transformations on uploaded audio. For example, texture/timbre morphing, style transfer, and effects-like processing. (2) **Large-model-driven generative music:** Using ComfyUI, we implement a text-to-music workflow with Stable Audio that supports prompt-conditioned control over style, tempo/BPM, instrumentation, and timbre palette, rapidly producing music clips with clear sectional structure. Especially, over the past two years, drawing on in-depth studies with more than 40 chefs and 20 diners, together with our work in the user community and design practice, we derive the following design considerations:

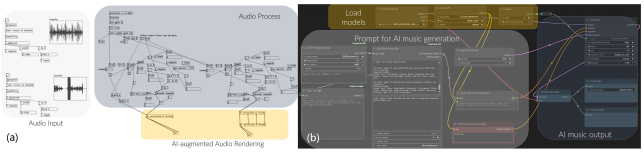

Figure 2: Ai music generation approach engagement. (a) The interface of pure data. (b) The interface of ComfyUI.

- ***DC1: Design chef-customized interfaces for creative agency***: Design the user interface specifically for chefs to move beyond a passive role in research and instead exercise creative agency. The interface should let them orchestrate music and system interactions in ways that align with their creative culinary thinking, supporting rapid exploration and iterative refinement.
- ***DC2: Design automated music generation aligned with culinary intent***: Chefs are not music-generation experts. Our experiments with two paradigms revealed a trade-off: the RAVE model affords improvised, performative experiences but its expressivity is bounded by internal parameters and input formats; Stable Audio, by contrast, enables direct, text-driven musical output that more closely captures chefs' intent. Combining these strengths suggests multi-modal workflows (e.g., jointly conditioning on audio, images, and text) to express implicit design ideas.
- ***DC3: Design diner-controlled interactions for dining engagement***: Although diners appreciated "sonic food," triggering fixed sound files solely through eating actions can reduce their perceived control. Drawing on real-time, responsive music generation from RAVE model, the design could benefit from a data-driven sonification approach in which diners' dining signals drive adaptive musical changes, enabling bidirectional interplay with chefs' compositions.
- ***DC4: Design instrumented tableware to strengthen the diner–food bond***: Create diner-specific interactive tableware that activates only upon meaningful contact with food,

preserving the inherent diner-food connection. Explore sensing designs and fabrication (e.g., 3D-printed components, hybrid sensors) that are food-safe, minimally intrusive, and robust in service environments.

3.2 Development of SoniFork System

SoniFork is a utensil equipped with capacitive sensing capabilities to detect diners' eating states. We designed the system to be as compact and universal as possible, mounting it on the end of a stainless steel spoon to avoid interfering with the eating process. The hardware includes an ESP32-S3, an Audio Breakout, a mini speaker, and a battery. We also developed a companion app that enables chefs to compose and orchestrate AI-generated music in real time (DC1), which can then be transmitted to SoniFork through WIFI to augment the dining experience (Figure 3). Additionally, variations in capacitance are mapped to dynamic sound representations through our custom sonification algorithm, creating real-time audio feedback that responds to diners' interactions (DC3).

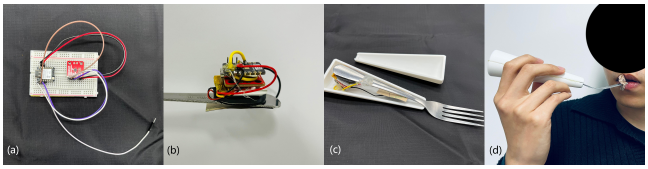

Figure 3: Prototyping process of SoniFork: (a) feasibility testing on a breadboard; (b) final prototype on a stainless-steel fork; (c) assembly within a 3D-printed enclosure; (d) a participant uses SoniFork to eat pickled ume, triggering audio feedback.

3.3 Sound Creation Workflow Design

Based on ComfyUI and the Stable Audio model, we designed four interaction workflows to represent chefs' culinary voices through AI-generated sound (DC2): non-input, visual input, audio input, and textual input. These modes reflect varying degrees of user control over the AI in this co-creative process (Section 2). The dynamic variation in agency afforded by these workflows offers diverse ways for chefs to express their culinary creativity, ranging from inspiration-driven exploration to precise control. Meanwhile, to accommodate the fact that chefs are not typically sound design experts, we also offer an optional integrated AI agent that polishes chefs' initial input related to their dishes into a more effective prompt (e.g., genre, BPM, mood, instrument types), thereby maximizing the generative model's performance.

4 Design Reflection and Application Scenarios

Chef–AI sound co-creation reduces barriers for chefs without formal music training, enabling personalized sonic expression of culinary intent. SoniFork enables sound exploration in non-specialist settings, offering a practical route for flavor adjustments and creative ideation. For example, an AI-driven image-to-music workflow maps a dish's visual features to auditory representations, constructing correspondences between culinary concepts and sound. Therefore, building on SoniFork, we outline three scenarios illustrating how chefs can use music to shape diners' experiences: (i) modulating the pace of eating, (ii) opening imaginative space, and (iii) influencing taste perception.

Scenario 1: Modulating Emotion through Musical Rhythm. Chefs can create musical rhythms aligned with distinct emotional states. For example, fast-paced and cheerful melodies may accompany lively meals, encouraging a sense of energy and social engagement, whereas slower and more melancholic tones may align with solitary or reflective dining contexts. By dynamically adjusting rhythmic structure, the system enable chefs to reshape diners' emotional experience, extending food consumption into an emotionally co-regulated activity between chefs and diners.

Scenario 2: Immersive Natural Soundscapes in Dining. The co-creation workflow in SoniFork also supports vivid environmental sound effects, enabling chefs to augment dishes with enhanced dining soundscapes that facilitate diners' dining association. For example, interactions can trigger birdsong, ocean waves, or rustling leaves, situating diners in an imagined natural setting [32]. By evoking environmental context, the system can promote relaxation and focus, thereby supporting mindful dining. This scenario highlights SoniFork's potential for experiential dining, transforming everyday eating into a multisensory and contextually immersive experience.

Scenario 3: Crossmodal Influence on Taste Perception. Leveraging crossmodal effects between sound and taste, chefs can use SoniFork to deliver auditory stimuli that modulate gustatory perception. In this application, sharp, high-pitched tones may be paired with foods to enhance perceived sourness, whereas low, resonant tones can intensify sweetness or a sense of richness. This approach expands the expressive latitude of culinary practice: taste is shaped not only during cooking but also dynamically evolved through the act of eating.

5 Conclusion and Limitations

In this paper, we presented the SoniFork system and its accompanying app, featuring four chef–AI sound co-creation workflows. This system empowers chefs to take ownership of sound design, allowing them to curate, shape, and generate dynamic soundscapes that evolve in response to diners' eating process. We articulated the design and implementation of this co-creative process. In future work, we plan to conduct a field study to investigate user experiences and derive design implications for creating AI-empowered interactive systems that enrich chefs' culinary creativity.

Acknowledgments

We sincerely thank all participants for their contributions and reviewers for their time and effort in fostering a better version of this paper. Florian 'Floyd' Mueller thanks the Australian Research Council, especially DP190102068, DP200102612, and LP210200656.

References

[1] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).

[2] Suno AI. 2024. *Suno AI.* https://suno.com/

[3] Gino Brunner, Andres Konrad, Yuyi Wang, and Roger Wattenhofer. 2018. MIDI-VAE: Modeling dynamics and instrumentation of music with applications to style transfer. *arXiv preprint arXiv:1809.07600* (2018).

[4] Clare Caldwell and Sally A Hibbert. 2002. The influence of music tempo and musical preference on restaurant patrons' behavior. *Psychology & Marketing* 19, 11 (2002), 895–917.

[5] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. " Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[6] 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*. 1–19.

[7] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and controllable music generation. *Advances in Neural Information Processing Systems* 36 (2023), 47704–47720.

[8] 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*. 1–19.

[9] 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

[10] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. 2020. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341* (2020).

[11] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable digital signal processing. *arXiv preprint arXiv:2001.04643* (2020).

[12] Rebecca Fiebrink and Baptiste Caramiaux. 2016. The machine learning algorithm as creative musical tool. *arXiv preprint arXiv:1611.00379* (2016).

[13] 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

[14] Alexandre RJ François, Elaine Chew, and Dennis Thurmond. 2007. Visual feedback in performer-machine interaction for musical improvisation. In *Proceedings of the 7th international conference on New interfaces for musical expression*. 277–280.

[15] Jules Françoise, Sarah Fdili Alaoui, and Yves Candau. 2022. CO/DA: Live-Coding Movement-Sound Interactions for Dance Improvisation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.

[16] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 937–946. doi:10.1145/2207676.2208538

[17] 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.

[18] Linjia He, Hongyue Wang, Sarah Goodwin, Benjamin Tag, and Don Samitha Elvitigala. 2024. Pedalling into the Future: Towards Enhancing Cycling Experience Using Augmented Reality. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 190–194. doi:10.1145/3656156.3663699

[19] Wen-Yi Hsiao, Jen-Yu Liu, Yin-Cheng Yeh, and Yi-Hsuan Yang. 2021. Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 178–186.

[20] Cheng-Zhi Anna Huang, Tim Cooijmans, Adam Roberts, Aaron Courville, and Douglas Eck. 2019. Counterpoint by convolution. *arXiv preprint arXiv:1903.07227* (2019).

[21] Yu-Siang Huang and Yi-Hsuan Yang. 2020. Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions. In *Proceedings of the 28th ACM international conference on multimedia*. 1180–1188.

[22] Hsiao-Tzu Hung, Joann Ching, Seungheon Doh, Nabin Kim, Juhan Nam, and Yi-Hsuan Yang. 2021. EMOPIA: A multi-modal pop piano dataset for emotion recognition and emotion-based music generation. *arXiv preprint arXiv:2108.01374* (2021).

[23] Purnima Kamath, Fabio Morreale, Priambudi Lintang Bagaskara, Yize Wei, and Suranga Nanayakkara. 2024. Sound designer-generative ai interactions: Towards designing creative support tools for professional sound designers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[24] Ryo Kanda, Mitsuyo Hashida, and Haruhiro Katayose. 2009. Mims: Interactive Multimedia Live Performance System. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 45–47.

[25] Yewon Kim, Sung-Ju Lee, and Chris Donahue. 2024. Amuse: Human-AI Collaborative Songwriting with Multimodal Inspirations. *arXiv preprint arXiv:2412.18940* (2024).

[26] Klemens Knöferle and Charles Spence. 2012. Crossmodal correspondences between sounds and tastes. *Psychonomic bulletin & review* (2012), 1–15. doi:10.3758/s13423-012-0321-z

[27] Dehui Kong, Hongyue Wang, Sijie Zhou, Hong Cui, and Zhiquan Feng. 2024. PenLab: Towards Understanding of Active Collaboration for Solid Geometry Teaching. 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 264, 7 pages. doi:10.1145/3613905.3650789

[28] Yu-Ren Lai, Hsi-Jen Chen, and Chia-Han Yang. 2023. Exploring the impact of generative artificial intelligence on the design process: Opportunities, challenges, and insights. *Artificial Intelligence, Social Computing and Wearable Technologies* 113 (2023), 49–59.

[29] Qingqin Liu, Ziqi Fang, Jiayi Wu, Shaoyu Cai, Jianhui Yan, Tiande Mo, Shuk Ching Chan, and Kening Zhu. 2025. VirCHEW Reality: On-Face Kinesthetic Feedback for Enhancing Food-Intake Experience in Virtual Reality. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '25)*. Association for Computing Machinery, New York, NY, USA, Article 19, 13 pages. doi:10.1145/3721238.3730694

[30] Xinyue Liu, Xipei Ren, Xinrui Ren, and Xiaoqing Sun. 2024. Hicclip: Sonification of Augmented Eating Sounds to Intervene Snacking Behaviors. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 1372–1384.

[31] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13. doi:10.1145/3313831.3376739

[32] Hong Luo, Pengtao Zhao, Hongyue Wang, Vincent van Rheden, Rakesh Patibanda, Yizhen Wang, Yuchen Hu, Don Samitha Elvitigala, and Florian 'Floyd' Mueller. 2025. *Respireal: Enriching Human-Nature Connection through a Breath-Controlled Mixed Reality Experience.* Association for Computing Machinery, New York, NY, USA, 351–356. https://doi.org/10.1145/3715668.3736338

[33] Marianna Obrist, Ferran Altarriba Bertran, Neharika Makam, Soh Kim, Christopher Dawes, Patrizia Marti, Maurizio Mancini, Eleonora Ceccaldi, Nandini Pasumarthy, 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

[34] Felipe Reinoso Carvalho, Abdellah Touhafi, Kris Steenhaut, Raymond van Ee, and Carlos Velasco. 2017. Using sound to enhance taste experiences: An overview. In *Bridging People and Sound: 12th International Symposium, CMMR 2016, São Paulo, Brazil, July 5–8, 2016, Revised Selected Papers 12*. Springer, 316–330. doi:10.1007/978-3-319-67738-5_19

[35] Charles Spence. 2015. Eating with our ears: assessing the importance of the sounds of consumption on our perception and enjoyment of multisensory flavour experiences. *Flavour* 4 (2015), 1–14. doi:10.1186/2044-7248-4-3

[36] Charles Spence, Felipe Reinoso-Carvalho, Carlos Velasco, and Qian Janice Wang. 2019. Extrinsic auditory contributions to food perception & consumer behaviour: An interdisciplinary review. *Multisensory research* 32, 4-5 (2019), 275–318.

[37] Lorenzo D Stafford, Mya Fernandes, and Ed Agobiani. 2012. Effects of noise and distraction on alcohol perception. *Food Quality and Preference* 24, 1 (2012), 218–224. doi:10.1016/j.foodqual.2011.10.012

[38] Udio. 2024. *Udio.* https://udio.com/

[39] Carlos Velasco, Marianna Obrist, Olivia Petit, and Charles Spence. 2018. Multisensory technology for flavor augmentation: a mini review. *Frontiers in psychology* 9 (2018), 26.

[40] Elena Vildjiounaite, Julia Kantorovitch, Vesa Kyllönen, Ilkka Niskanen, Mika Hillukkala, Kimmo Virtanen, Olli Vuorinen, Satu-Marja Mäkelä, Tommi Keränen, Johannes Peltola, et al. 2011. Designing socially acceptable multimodal interaction in cooking assistants. In *Proceedings of the 16th international conference on Intelligent user interfaces*. 415–418.

[41] 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. *Proceedings of the 2025 CHI conference on human factors in computing systems* (2025), 1–19. doi:10.1145/3706598.3714237

[42] Hongyue Wang, Jialin Deng, Dehui Kong, Ziqi Fang, Hong Luo, Nandini Pasumarthy, Rakesh Patibanda, Sasindu Abewickrema, Xiao Zoe Fang, Don Samitha Elvitigala, et al. 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*. 1–8. doi:10.1145/3706599.3719700

[43] 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

[44] Hongyue Wang, Zhiquan Feng, and Xin Meng. 2023. Multimodal information perception and understanding: application of smart glove in virtual-reality fusion chemistry experiment platform. *Advanced Materials Technologies* 8, 10 (2023), 2200549. doi:10.1002/admt.202200549

[45] Hongyue Wang, Zhiquan 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

[46] Hongyue Wang, Zhiquan 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

[47] Hongyue Wang, Xin Meng, and Zhiquan Feng. 2021. Research on the Structure and Key Algorithms of Smart Gloves Oriented to Middle School Experimental Scene Perception. In *CCF Conference on Computer Supported Cooperative Work and Social Computing*. Springer, 409–423. doi:10.1007/978-981-19-4546-5_32

[48] Yan Wang, Zhuying Li, Robert S Jarvis, Joseph La Delfa, Rohit Ashok Khot, and Florian Floyd Mueller. 2020. WeScream! toward understanding the design of playful social gustosonic experiences with ice cream. In *Proceedings of the 2020 ACM designing interactive systems conference*. 951–963. doi:10.1145/3357236.3395456

[49] Yan Wang, Zhuying Li, Robert S Jarvis, Angelina Russo, Rohit Ashok Khot, and Florian'Floyd' Mueller. 2019. Towards understanding the design of playful gustosonic experiences with ice cream. In *Proceedings of the annual symposium on computer-human interaction in play*. 239–251.

[50] Danielle Wilde and Ferran Altarriba Bertran. 2019. Participatory Research through Gastronomy Design: A designerly move towards more playful gastronomy. *International Journal of Food Design* 4, 1 (2019), 3–37. doi:10.1386/ijfd.4.1.3_1

[51] 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*. 1–7.

[52] Liran Zhou, Zhiquan Feng, Hongyue Wang, and Qingbei Guo. 2024. MIUIC: a human-computer collaborative multimodal intention-understanding algorithm incorporating comfort analysis. *International Journal of Human–Computer Interaction* 40, 19 (2024), 6077–6090. doi:10.1080/10447318.2023.2247606

A Appendix: Description of the Prompting Framework

You are an AI-powered assistant designed to help chefs generate audio that complements their culinary creations. Chefs will provide abstract descriptions of their dishes, ingredients, and inspirations, and your role is to translate these descriptions into simple and structured prompts optimized for Stable Audio's AI model. Your only task is to convert the chef's dish description into a concise, structured prompt following a fixed format.

Instruction Guidelines: 1. Understanding Chef Inputs Chefs will describe their dish using ingredients, cooking techniques, cultural references, moods, and themes. Inputs may be abstract, such as "warm and comforting like a grandma's soup" or "sharp and edgy like a spicy Thai curry." Your task is to analyze their intent and map it to an appropriate audio style.

2. Strictly follow this format for output: Format: [Band/Solo] | Genre: [Main Genre] | Subgenre: [Optional] | Instruments: [Key Instruments] | Moods: [Emotions] | BPM: [Tempo] | Style: [Ambiance] Do not include any additional details, explanations, or breakdowns. Keep it concise—each field should contain no more than 3-5 words. Omit the "Atmosphere" field—only use the required structured format.

3. Converting Inputs into Audio Descriptors For each chef's description, extract key elements and structure them into a well-defined prompt format that includes: Genre: (e.g., Jazz, Lo-fi, Orchestral, Ambient, Nature Soundscape, Traditional Folk) Sub-genre: (if applicable, e.g., "Minimalist Piano," "Ethereal Ambient," "Tribal Drumming") Instruments: (e.g., Soft piano, deep bass, ethnic percussion, strings, synthesizers) Moods: (e.g., Warm, Nostalgic, Exciting, Mysterious, Playful) BPM (Tempo): (e.g., Slow for a relaxing dish, Upbeat for energetic foods) Styles: (e.g., "Fine Dining Elegance," "Street Food Vibes," "Homecooked Comfort") Atmosphere: (e.g., "Rainy evening in a small Parisian café" for a French dish)

4. Refinement and Enhancement (1)Expand vague descriptions from the chef (e.g., "sweet" → "warm, soft piano melody"). (2)Ensure a well-rounded prompt covering music genre, instruments, and emotional tone. (3)Adjust BPM to match the intended pace of the dining experience. (4)If input is too simple, add enriching details (e.g., environmental sounds, specific instruments).

5. Example: (1) Chef's Input: "A bold and fiery Mexican taco with smoky chipotle and fresh lime." Generated Prompt: Genre: Latin | Subgenre: Mexican Percussion Groove | Instruments: Congas, Acoustic Guitar, Brass Stabs | Moods: Bold, Vibrant, Lively | BPM: 120 | Style: Street Food Energy

(2) Chef's Input: "A delicate truffle risotto, creamy and luxurious, with deep umami flavors." Generated Prompt: Format: Band | Genre: Orchestral Jazz | Instruments: Soft piano, upright bass, smooth saxophone | Moods: Elegant, Warm, Rich | BPM: 85 | Style: Fine Dining Sophistication

(3) Chef's Input: "A slow-cooked beef stew, hearty and comforting, like a warm hug." Generated Prompt: Format: Solo | Instruments: Acoustic Guitar, Harmonica, Soft Drum Brushes | Moods: Cozy, Nostalgic, Earthy | BPM: 70 | Style: Country Folk

(4) Chef's Input: "A deconstructed dessert—unexpected, playful, and futuristic." Generated Prompt: Format: Band | Genre: Experimental Electronica | Instruments: Glitch Synths, Digital Harp, Reversed Strings | Moods: Playful, Mysterious, Evolving | BPM: 110 | Style: Avant-Garde