2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3503-2838-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR59233.2023.00106

Effects of visual presentation near the mouth on cross-modal effects of multisensory flavor perception and ease of eating

Kizashi Nakano* The University of Tokyo
Monica Perusquia-Hernandez† Nara Institute of Science and Technology
Naoya Isoyama‡ Otsuma Women's University
Hideaki Uchiyama§ Nara Institute of Science and Technology
Kiyoshi Kiyokawa¶ Nara Institute of Science and Technology

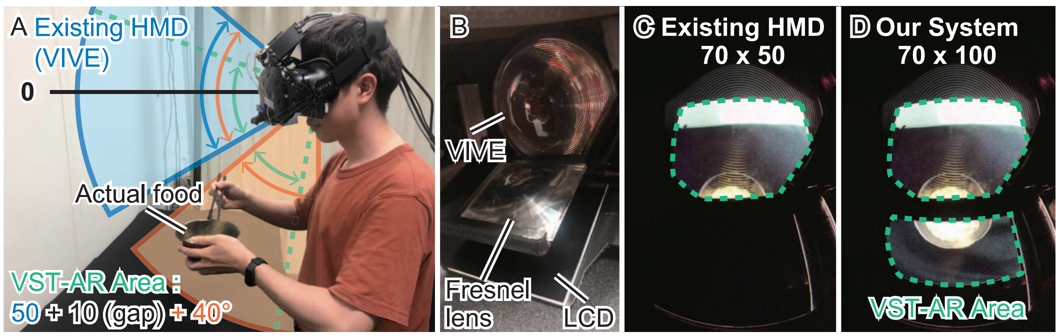

Figure 1: A) Side view. Blue: vertical FoV of the HTC VIVE (approx. 70 [deg]). Part of the lower FoV is missing owing to the LCDs housing (approx. 10 [deg]). Orange: increased vertical FoV (approx. 60 [deg]). Green: vertical FoV of a VST–AR (approx. 100 [deg]). B) The added LCD and Fresnel lens. C) Original FoV of the HTC VIVE (approx. 70×50 [deg]). Only the top half of the actual food is visible. D) FoV of the prototype (approx. 70×100 [deg]). The upper and lower parts of the actual food are visible at the same time.

Abstract

Various studies have suggested that altering the appearance of food can impact multisensory flavor perception. The cross-modal effect of such visual changes on gustation may allow for the presentation of food tastes that are difficult to express with simple combinations of taste stimuli. This cross-modal effect of visual changes on gustation holds potential for applications in gustatory displays. However, the current limitation of existing Head-Mounted Displays (HMDs) is their restricted vertical Field of View (FoV), which prohibits the display of images near the mouth while eating. This limitation may impede the cross-modal effect of visual changes on multisensory flavor perception. Additionally, the lack of visibility around the mouth area challenges the ease of eating. To address these issues, we design a Video See-Through (VST)-HMD with an expanded vertical FoV (approx. 100 [deg]). Using the HMD, we investigated how presenting visual information near the mouth affects the cross-modal effects of flavor perception and ease of eating. In our experiment, machine learning techniques were utilized to alter the appearance of food. However, the result showed no significant differences in the amount of cross-modal effects or the ease of eating between the groups with and without visual information near the mouth. As a discussion of this result, the participants may not direct their visual attention to the food when they put the food in their mouths. The experiment also examined whether visual changes alter the taste as well as the smell and texture of the food. The findings demonstrated that visual changes could present the smell and texture of the food following the modifications. This result was confirmed irrespective of the visibility near the mouth.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 Introduction

The question of how to create safe and flavorful food has been a long-standing research topic for humans. However, limited engineering methods are available to manipulate people's flavor perception and make them perceive a specific taste or food type. One of the reasons for this is the difficulty in defining a fundamental basis for gustatory perception, similar to the three primary colors for vision [6]. Gustatory perception is not solely driven by pure gustatory stimuli. Instead, it is influenced by the integration and interaction of multiple senses, including vision, hearing, olfaction, and tactile perception, which work together to perceive taste [26, 32]. In this paper, we define multisensory flavor perception [32] as the taste we perceive by integrating our various senses.

In light of these challenges, research has focused on presenting gustatory experiences by leveraging multisensory integration and modifying stimuli from other senses rather than relying solely on taste stimuli [12, 18, 19, 22]. For instance, Nakano et al. developed a gustatory manipulation interface using cross-modal effects to alter the appearance of food through Head Mounted Displays (HMD) and Augmented Reality (AR) technology [18, 19]. They successfully induced the perception of specific tastes using their gustatory manipulation interface based on techniques such as StarGAN [3]. In addition, studies exploring the application of Mixed Reality (MR) and Virtual Reality (VR) technology in the dining domain have been on the rise and are attracting increasing attention [1]. These studies investigate the effects of the visual appearance of Virtual Environments (VEs) on gustation [2, 24, 25, 34]; the impact of eating with virtual friends in a VE [13]; and the development of software tools

*e-mail: kizashi@cyber.t.u-tokyo.ac.jp
†m.perusquia@is.naist.jp
‡isoyama@otsuma.ac.jp
§hideaki.uchiyama@is.naist.jp
¶kiyo@is.naist.jp

DOI 10.1109/ISMAR59233.2023.00106

to support eating experiences in virtual environments [17]. These endeavors expand our understanding of how sensory cues can be manipulated in virtual dining scenarios and pave the way for innovative approaches to enhance the overall dining experience.

However, conventional HMDs for VR have a limited field of view (FoV) compared to human vision. Typically, they offer a FoV of approximately 90–110 [deg] in the diagonal direction and 70–90 [deg] in the vertical direction. The vertical FoV of human vision is approximately 120–135 [deg], with the downward FoV below the horizontal line of sight spanning approximately 70–80 [deg], larger than the upward FoV of around 50–55 [deg], whereas conventional HMDs are covered only about 50% of humans' downward FoV. As a result, it is currently not feasible to display food near the mouth area while wearing the HMD.

This narrower FoV, particularly in the vertical direction, poses a challenge when displaying food near the user's mouth while wearing the HMD. The limited FoV of HMDs creates two significant problems. The first problem relates to the ease of eating when wearing the HMD. With a clear view of the mouth area, users find it easier to accurately determine the relative position of the food in relation to their mouth. The front camera of an HMD, which captures images of the food, is usually positioned across the entire device surface. Consequently, there is a disparity between the camera's position and the actual position of the user's eyes. Furthermore, as the human face is structured such that the mouth is directly below the eyes, users may mistakenly perceive the mouth as directly beneath the front camera position [17–19]. Therefore, displaying food near the mouth is impossible while wearing the HMD.

The second problem arises from the inability to alter the appearance of food near the mouth while wearing a conventional HMD. Previous studies demonstrated that multisensory flavor perception could be altered through a cross-modal effect of visual modulation, which changes the appearance of food [18, 19]. However, displaying the altered food appearance at the precise moment when the user perceives the taste, such as when the food is placed in the mouth, is currently not feasible. To explore the impact of food image presentation near the mouth, an HMD with an expanded vertical viewing angle is necessary. Researchers have proposed enhancing the vertical FoV in VR-HMDs by incorporating an additional optical system [5, 20]. However, implementing a Video See-Through (VST)-AR function to complement such a wide viewing angle is yet to be realized. In this study, we developed an AR system that can present visual information near the mouth by adding a VST function to the VR-HMD with a larger downward FoV.

We hypothesize that the limited downward FoV suppressed the cross-modal effects of visual modulation on gustation and investigated the effects of downward FoV expansion on the ease of eating and cross-modal effects. Furthermore, we expanded our investigation beyond multisensory flavor perception and explored the effects of visual modulation on other sensory aspects, such as smell and food texture. This approach was motivated by previous studies [18, 19] that suggested the potential influence of visual changes on multiple sensory dimensions. Hence, we sought to determine whether visual modulation could alter the perception of smell and food texture.

The major contributions of this study are as follows:

- We developed a novel VST-HMD with additional display and fisheye camera units for a wide downward FoV.
- We report that in the experiment with the downward FoV expansion, significant differences were not found in the ease of eating while wearing an HMD and the amount of gustatory manipulation of the cross-modal effect.
- We demonstrate that gustatory manipulation by visual modulation can change the smell and food texture from the original food to the intended food.

2 Related Work

2.1 Effects of visual modulation on the taste of food

The first method we can use to describe the taste of food in an engineering way is to present a combination of basic tastes for the sense of taste (gustatory sensations). However, there are several definitions of basic tastes, such as the research report that taste is a continuum and that there is no such thing as a basic taste [6]; and the research report that there is no single receptor site for sweet and bitter tastes [4]. In addition, a method for presenting taste has yet to be established due to the influence of painful stimuli such as pungency and the fact that the threshold for perceiving taste (e.g., sweetness) differs among individuals [11].

Alternatively, research has revealed that our gustatory sensations are affected by gustatory stimuli and by the other senses [32]. Ernst et al. proposed a multisensory integration model using Bayes' theorem to explain the nature of multisensory integration [7, 8]. Research has been conducted to induce a cross-modal effect through multisensory presentation by utilizing the integrated cognitive characteristics of humans and presenting information from various senses to present gustation.

Indeed, research has demonstrated that vision plays a significant role in influencing gustation [33]. Several studies have successfully manipulated multisensory flavor perception by altering its color [16, 23, 28]. For instance, Morrot et al. showed that changing the color of wine could lead sommeliers to perceive a different taste, making white wine taste like red wine [16]. Narumi et al. employed LEDs to color beverages and demonstrated that they could change the perceived flavor of the same beverage to a different flavor [23]. Similarly, Ranasinghe et al. combined color, smell, and electrical taste stimuli to alter the perceived flavor of water in a cocktail glass [28]. These findings collectively support the idea that visual cues, such as color, can influence how we perceive the flavor of food and beverages. Researchers have explored various approaches to enhancing the gustatory experience by leveraging the cross-modal effects of vision on gustation.

AR using HMD technologies can modify not only the basic color attributes but also the visual presentation of the food itself. For instance, one can alter the flavor perceived by overlaying a three-dimensional model or image onto the food [18, 19, 22, 38]. By combining an olfactory display with various cookie images on top of plain cookies, Narumi et al. could transform the plain cookies into many different perceived flavors, for instance, chocolate [22]. Ueda and Okajima developed a gustatory manipulation interface that uses machine learning to detect tuna sushi and change its appearance to salmon sushi, flatfish sushi, medium fatty tuna, and the fattiest portion of tuna [38]. As a result, participants perceived more mouthfeel and oiliness in medium-fatty tuna and the fattiest portion of tuna sushi than tuna sushi. Nakano et al. developed a gustatory display that changed the appearance of food into the appearance of different foods and succeeded in changing the taste of actual somen noodles into the taste of fried noodles and ramen noodles [19]. In this work, food images were generated by a StarGAN [3]. They also investigated the persistence of the cross-modal effect of such visual changes on taste. They reported that the change in taste persisted until the fifth bite of the meal while viewing the appearance of the food after the change [18].

However, the current research on gustatory displays using HMDs faces a limitation in presenting images close to the mouth, mainly due to the restricted vertical viewing angle of the HMD. This constraint hinders altering the visual appearance of the food precisely at the moment of multisensory flavor perception. Furthermore, studies have indicated that wearing an HMD during eating can adversely affect the overall ease of eating [17–19]. Consequently, the objective of this study is to explore the potential of video presentation near the mouth area in enhancing cross-modal effects and improving the ease of eating.

2.2 Effect of visual modulation on smell and food texture

Olfaction exerts a profound influence on the multisensory perception of flavor [31]. For example, food flavor is heavily affected by signals from the taste buds associated with smell [30, 31, 35]. Stevenson et al. demonstrated that sucrose's sweet taste was enhanced while citric acid's sour taste was diminished when participants were exposed to the smell of sweet caramel [35]. They have indicated that when the olfactory system is blocked, up to 80% of multisensory flavor perception is diminished. In the context of gustatory displays, researchers have used cross-modal effects with olfactory displays to manipulate multisensory flavor perception [22, 28]. These findings highlight the significance of incorporating olfactory stimuli in various gustatory display techniques.

In addition, the tactile and auditory sensations associated with food texture play a crucial role in gustatory displays [10, 12]. Iwata et al. have devised a food simulator as a gustatory manipulation interface that can replicate food texture by capturing and reproducing the temporal changes in force during chewing [10], enabling users to experience the tactile aspects of different food textures. Furthermore, the influence of auditory stimuli on multisensory flavor perception has been investigated. Zampini and Spence demonstrated that when participants were exposed to high-pass filtered chewing sounds or white noise while consuming deep-fried potato chips, the chips were perceived as crispier and fresher [42]. Similarly, Koizumi et al. successfully enhanced the perceived crunchiness of potato chips, the thickness of cookies, and the stickiness of daifuku by processing the subject's chewing sounds and playing them back through earphones [12].

As described above, studies have explored gustatory displays that leverage the cross-modal effects of smell and food texture on multisensory flavor perception. Alternatively, when gustatory manipulation using visual modulation in gustatory displays is successful, participants often report changes not only in taste but also in the perceived smell and food texture of the food [18, 19]. To investigate how gustatory manipulation by visual modulation changes the olfaction and tactile, we added questions about the smell and food texture to the questions in the previous study [18, 19] in this experiment to investigate quantitatively.

2.3 HMD with wide viewing angle

HMDs with an expansive FoV are recognized for their ability to enhance the immersive spatial experience, despite the associated risk of inducing cybernetic disorientation [14]. As a result, there is a growing interest in the research and development of HMDs with wide FoVs, exemplified by HMDs such as the StarVR One [1].

Ratcliff et al. introduced a novel approach incorporating a contoured display coupled with a corresponding microlens array to achieve an HMD with an extensive FoV [29]. Panasonic has made a wide horizontal FoV through a dual-display system for each eye, seamlessly integrated with two lenses [2]. Rakkolainen et al. have developed a solution to expand the FoV of HMDs by incorporating additional lenses and displays to enhance the horizontal peripheral vision [27]. However, conventional wide FoV HMDs primarily focus on front-facing and horizontal image rendering, which presents a challenge when displaying content within the downward FoV.

Specific academic investigations have proposed methodologies for pseudo-expansion of the FoV [40, 41], predicated on the phenomenon of peripheral vision's diminished sensitivity to information relative to the central visual field [9]. Xiao et al. conceived a method of augmenting the viewing angle by arraying a multitude of full-color Liquid Crystal Displays (LCDs) and diffuser plates exterior to the eyepiece optics of the HMD, synchronizing them impeccably with the primary display of the device [40]. Yamada et al. employed a combination of a convex lens with standard magnification and a high-magnification Fresnel lens to project a slightly blurred image through the Fresnel lens within the peripheral FoV, thereby achieving a cost-effective approach to expand the viewing angle [41]. These approaches, albeit facilitating the expansion of the downward FoV, concurrently compromise the resolution of images within the peripheral visual field.

Lindeman et al. devised a system that enables the visualization of the surrounding physical environment within the downward FoV by incorporating an aperture below the eyepiece lens, equipped with an adjustable transparency liquid crystal shutter [15, 37]. While this system allows for the perception of the real environment in the downward FoV, it does not support image display. Endo et al. introduced a head-mounted display that expands the horizontal and vertical FoV by incorporating smartphones on the left, right, and bottom sections of the device [5]. However, the practicality of eating while using this HMD is compromised due to the added smartphones' complete coverage of the mouth area.

As previously discussed, most research efforts in expanding the FoV of HMDs have primarily concentrated on widening the horizontal viewing angle. In contrast, the vertical viewing angle has received less attention. Consequently, the presentation of visual information in the vicinity of the mouth, which is essential for gustatory manipulation through visual modulation, has been limited. In response to this challenge, Nakano et al. developed a VR-HMD that enhances the downward FoV by integrating two sets of lenses and displays oriented towards the vertical downward direction, thereby addressing this limitation [20]. However, this VR-HMD lacks a VST-AR functionality, precluding food image display.

In this study, we developed a VST-HMD with an enlarged downward FoV that can change the appearance of food in the real world using an improved HMD with an enlarged downward FoV developed. Also, we added a VST-AR function, enabling visual information to be presented near the mouth.

3 VST-HMD THAT EXPANDS THE DOWNWARD FOV

3.1 Hardware system configuration

We have designed an HMD that enhances the downward FoV in the VST area, enabling the presentation of visual information in that particular region (Figure 1). Schematic and physical representations of the proposed device are depicted in Figures 2 and 3, respectively. This HMD was improved from previous research [17] by adding VST functionality based on the VR-HMD with expanded downward FoV. For our augmented HMD, the external components of the HTC VIVE were removed, and two sets of an LCD module (Sharp LS029B3SX02; dimensions: 2.9 [inches]; display area: 51.84×51.84 [mm]; resolution: 1440×1440; frame rates: 120 [fps]) and a Fresnel lens (length: 35 [mm]; width: 38 [mm]; focal length: 40 [mm]) were positioned at an angle of 20 [deg] beneath each eyepiece within a three-dimensional (3D) printed housing. The eye relief and the distance from the lens to the LCD were set at 28 [mm]. Although this distance is shorter than the lens's focal length, the displayed image remains clear. The distance between the user's eyes and the HTC VIVE eyepiece was adjusted to approximately 13 [mm] at its maximum extent to accommodate the additional display units. In order to provide visual modulation to foods near the mouth, a fisheye camera (ELP-SUSB1080P01-L170; resolution, 1080×1920; FoV, 170 [deg]; frame rate, 50[fps]) was added to the front of our HMD prototype, tilted 25 [deg] downward from the front.

The actual FoVs of VST measured by a pre-calibrated equidistant projection fisheye camera (GS-15WDCM-1.5MM; resolution, 1920×1080; FoV, 180 [deg]) are shown in Table 1. As indicated in 1, implemented prototype significantly enhances the downward VST-FoV by approximately 50 [deg] ($10 + 40$ [deg]). Figure 1 (C, D) presents a VST-HMD prototype displaying real-life environmental

[1] StarVR, `https://www.starvr.com/`, last accessed August 22, 2023.
[2] Panasonic, `https://channel.panasonic.com/contents/19737/`, last accessed August 22, 2023.

images. In Figure 1 (C), which utilizes the original viewing angle of the HTC VIVE, only the upper portion of the food is visible. However, in Figure 1 (D), the entire food item can be displayed with an expanded lower viewing angle. The 3D printed housing results in a reduction of approximately 10 [deg] in the downward FoV of the original HTC VIVE. Consequently, the food in the figure appears to be divided by this gap. This representation aims to demonstrate how the food is displayed on both the upper and lower displays within the FoV. In practical usage, when the user consumes food using an HMD with an enlarged downward FoV, the food is often positioned within their reach. As a result, most of the food can be observed within the expanded downward FoV highlighted in orange in Figure 1 (A). In the future, dedicated fused lenses could be employed to mitigate the reduction of the FoV [2].

The vertical viewing angle of the HMD without the VST display is 130 [deg] (70 + 10 (gap) + 50 [deg]), which aligns with the findings of the previous study [17], as depicted by the blue and orange colors in Figure 1 (A). In the case of the VST display, the vertical viewing angle is 100 [deg] (50 + 10 (gap) + 40 [deg]), as indicated by the green line in Figure 1 (A). The reduction in the vertical viewing angle is attributed to the distortion-corrected fisheye camera used in this study (with a viewing angle of 170 [deg]) not fully encompassing the vertical viewing angle required for the downward-focused HMD. Thus, expanding the VST display area can be achieved by employing a fisheye camera with a wider viewing angle or combining multiple webcams. It is noteworthy that the vertical human FoV typically ranges from approximately 120 to 135 [deg], with the downward FoV spanning about 70 to 80 [deg]. The newly developed downward-field-expanding HMD features a downward FoV of 70 [deg], which sufficiently covers the downward FoV of humans.

The HTC VIVE officially announced by HTC has an FoV of approximately 110 [deg] in the horizontal, vertical, and diagonal directions, as indicated by their circular viewport. However, according to reports by iNFINITE, the actual FoV of the HTC VIVE is approximately 89 [deg] in both the horizontal and vertical directions [3]. The slight decrease in the FoV of the HTC VIVE from its official specification may be attributed to the larger distance between the eyepiece of the HTC VIVE and the user's eye, which was increased by approximately 13 [mm] in our specific setup.

3.2 Software system configuration

The software for the proposed system was developed using Unity 2019.4.39f1. Within the software, two virtual cameras were employed to project images onto the LCDs, aligning with the main camera of SteamVR. To ensure accurate image presentation, the virtual cameras' initial intrinsic and extrinsic parameters were determined by taking into account the physical dimensions of the display units. The edited view frustum was utilized to correct the image output from the virtual sub-camera. Additionally, the distortion induced by the Fresnel lens was manually rectified through mesh deformation, as depicted in Figure 4.

In the system of previous study [17], the initial intrinsic and extrinsic parameters of the virtual camera were only fine-tuned manually. As a result, straight lines were distorted, as shown in Figure 4 (A1). In addition, there was a problem that the image was enlarged compared to that shown in Figure 4 (B1), which is an image that can be originally displayed. In this study, by deforming the images using the accurate images taken in the real world shown in Figure 4 (C), we could display images closer to how they were initially seen.

We checked the stereo visibility with several individuals. We fixed the IPD of the VIVE and that of the lower optics to 69.4 [mm] and 70 [mm], respectively, as the optimal common distance.

[3]iNFINITE, https://www.infinite.cz/projects/HMD-tester-virtual-reality-headset-database-utility, last accessed August 22, 2023.

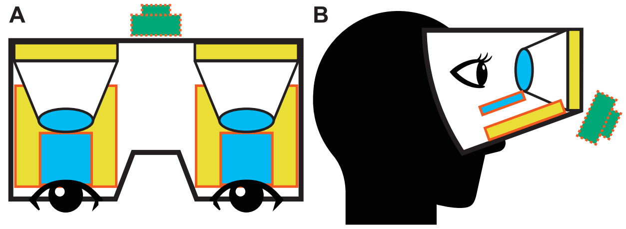

Figure 2: Schematic of the proposed HMD. A) Top view, B) Side view: yellow, display units; light blue, eyepiece; light green, fisheye camera; black lines, exterior of the HTC VIVE; orange lines, newly added parts.

Figure 3: Prototype HMD. A) Front view, B) Back view. The white part of the right eye is the sponge between the face and the HMD. The sponge in the left eye area is removed for the photo, C) Bottom display placement relative to the face.

3.3 Visual modulation function using machine learning

To modify the appearance of food, we utilized a system based on StarGAN [3], as previously studied [19]. The system comprises a client developed using Unity 2019.4.39f1 and a server created using the Python framework Flask. Initially, the server captures images from the fisheye camera attached to the front of the HMD. Subsequently, the fisheye camera distortion is corrected using OpenCV, and a visually modified food image is generated utilizing StarGAN. The resolution of the fisheye camera was limited to 480×640 to reduce delay. Consequently, the resulting resolution of the food image was 256×341, accounting for the distortion correction of the fisheye camera and subsequent resizing for image generation. The measured delay of the system was approximately 200 [ms]. Finally, the client displayed the generated food image at a distance of 1 [m] from the user's eyes, tilted downward by 25 [deg], emulating the perspective of a fisheye camera.

4 EXPERIMENT

4.1 Overview

In this experiment, our objective was to explore the potential impact of an expanded downward FoV on the visibility of food near the mouth and its subsequent influence on gustatory manipulation through visual modulation. Additionally, we aimed to assess the ease of eating food while wearing the HMD under these conditions. Furthermore, we investigated the cross-modal effect of perceiving the smell and texture of food through visual modulation, as previously reported in a previous study [19].

We utilized cups of somen noodles as the actual food, the same food used in the previous studies [18, 19]. It is important to note

Table 1: FoV of the VST area, measured with an equidistant projection fisheye camera. FoV is shown as a range because a portion of the VST image is not displayed in the added downward viewing area.

FoV	Horizontal	Vertical (upper + gap + lower)
HTC VIVE	∼70 [deg]	∼50 (20 + 0 + 30) [deg]
Our prototype	∼(50–70) [deg]	∼100 (20 + 10 + (55–70)) [deg]
Difference	±0 [deg]	+(35–50) (0 + 10 + (25–40)) [deg]

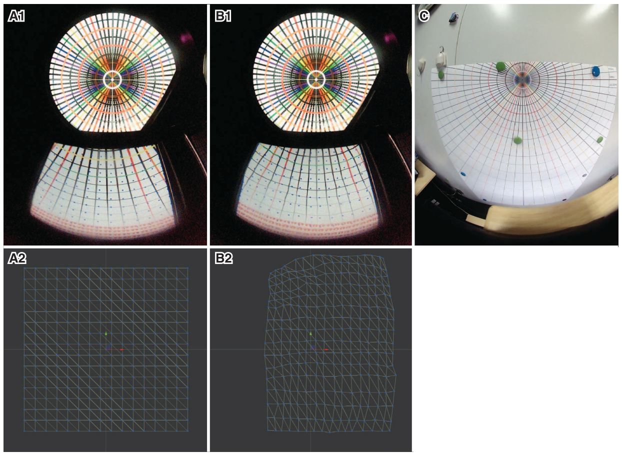

Figure 4: Images displayed on the HMD and the mesh shape before and after mesh deformation. A1) Display image before mesh deformation captured by the fisheye camera. A2) Mesh shape before deformation. B1) Display image after mesh deformation captured by the fisheye camera. B2) Mesh shape after deformation. C) Calibration pattern placed in front of the fisheye camera; each line represents a viewing angle in 5 [deg] increments.

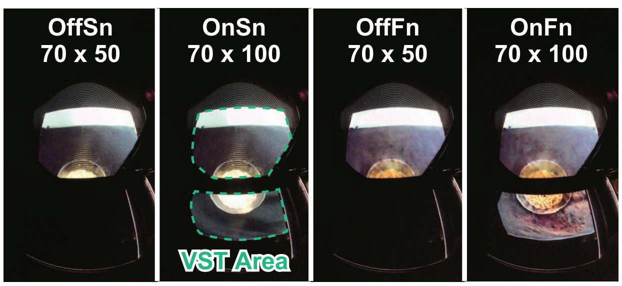

Figure 5: Viewing angle and video output for each condition. OffSn, OffFn) horizontal: approx. 70 [deg], vertical: approx. 50 [deg]. OnSn, OnFn) horizontal: approx. 70 [deg], vertical: approx. 100 [deg] (maximum viewing angle displayed))

that only the soup powder was used, and no condiments were added. The experimental design followed a 2×2 factorial design. The independent variables included the downward FoV (**O**N: additional display units were utilized; **O**FF: the units were not used) and the food appearance (**S**n: original food without visual modulation, i.e., somen noodles; **F**n: transformed food with visual modulation, i.e., fried noodles). Consequently, a total of four conditions were investigated: OffSn (no additional display units, somen noodles), OnSn (additional display units, somen noodles), OffFn (no additional display units, fried noodles), and OnFn (additional display units, fried noodles).

The visual depiction of the food appearance under each condition is presented in Figure 5. All variables were within-participant. Participants performed the tasks in a counterbalanced order using a Latin square design. This experiment's evaluation encompassed sensory evaluations conducted through a questionnaire and quantitative assessment based on head angle and elapsed time during eating.

4.2 Procedure

The experiment was conducted in a quiet room in our laboratory. The room was furnished with a black desk and chair facing the white wall, with red marks drawn on them so that they appear to be in front of them when they sit down.Before the experiment, participants were instructed to ensure they were neither moderately hungry nor full. Upon sitting at the desk, they were informed about the nature of the experiment, which involved consuming somen noodles under different conditions. Participants were also provided with a questionnaire, and they confirmed the content. The experiment proceeded in the following sequential order:

1. Participants drink water to wash out any food remaining in their mouths before starting each condition.
2. They wear HMDs with increased downward FoV.
3. They hold the bowl and chopsticks in their hands for 10 [sec] in each condition and visually inspect the somen noodles in the bowl.
4. They place the bowl and chopsticks on the table and adjust their position so that the red markings on the wall are visible in front of them.
5. They are told to begin the task.
6. They eat more than two bites of somen noodles using chopsticks.
7. After visually confirming the red markings on the wall in front of them, participants announce the end of the task.
8. They remove the HMD and proceed to fill out a questionnaire.
9. Steps 1–8 are repeated for each condition.

Step 3 was conducted to assess the visual presentation of the food under each condition. The deliberate movement to observe the red marks in steps 4 and 7 was executed to standardize the initial viewing angles for accurate pitch angle measurements. The duration of the experiment was measured during the interval spanning from steps 5 to 7. It is important to note that the food was only shown to participants while they are wearing the HMD to ensure that the experimental results are not influenced by the original food appearance.

4.3 Questionnaire

The questionnaire comprised eleven inquiries to assess taste, type, appearance, smell, texture, and ease of eating. These questions were designed to explore the impact of expanding the downward FoV on multisensory flavor perception.

The specific questions are as follows:

Q1. It tasted like somen noodles.
Q2. It tasted like fried noodles.
Q3. It felt like I was eating somen noodles.
Q4. It felt like I was eating fried noodles.
Q5. It felt the appearance of food was somen noodles.
Q6. It felt the appearance of food was fried noodles.
Q7. It felt the smell of somen noodles.
Q8. It felt the smell of fried noodles.
Q9. It felt the food texture of somen noodles.
Q10. It felt the food texture of fried noodles.
Q11. It was easy to eat with the VR goggle (HMD).

Q1 to **Q**4 were used to measure perceived taste and recognized type of food similar to previous studies [17, 19] to investigate the effect of the downward FoV expansion on multisensory flavor perception. **Q**5 and **Q**6 were used to measure whether visually presented food images appeared to be the food we intended. **Q**7 to **Q**10 were used to measure whether the cross-modal effects of visual changes affected olfaction and tactile. **Q**11 was used to measure the ease of eating. All questions were asked in random order on a scale of 101 using the VAS [0 (strongly disagree) to 100 (strongly agree)].

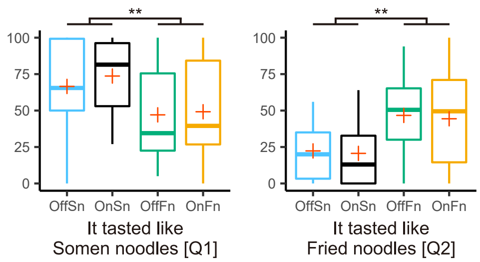

Figure 6: The perceived taste of food results (Q1, Q2) [0 (strongly disagree) to 100 (strongly agree)].

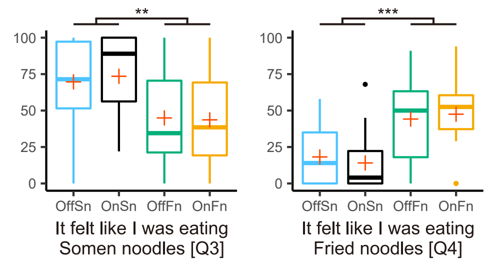

Figure 7: The recognized type of food results (Q3, Q4) [0 (strongly disagree) to 100 (strongly agree)].

4.4 Hypotheses

In this experiment, we set the following hypotheses.

- **H**1 Increasing the downward FoV enhances the cross-modal effect on gustation by improving the visibility of visually modulated food. This, in turn, yields an interaction in the results of **Q**1, **Q**2, **Q**3, and **Q**4.
- **H**2 Increasing the downward FoV improves the visibility of the food after visual modulation, which is felt in the appearance of the visually presented food, which in turn found an interaction in the results of **Q**5 and **Q**6.
- **H**3 Cross-modal effects with visual modulation evoke the smell of the presented food, which in turn decreases **Q**7 scores and increases **Q**8 scores in the **F**n condition.
- **H**4 Cross-modal effects with visual modulation evoke the food texture of the presented food, which in turn decreases **Q**9 scores and increases **Q**10 scores in the **F**n condition.
- **H**5 Increasing the downward FoV enhances the cross-modal effect on olfaction, which in turn finds an interaction in the results of **Q**7 and **Q**8.
- **H**6 Increasing the downward FoV enhances the cross-modal effect on tactile, which in turn finds an interaction in the results of **Q**9 and **Q**10.
- **H**7 Increasing the downward FoV improves the ease of eating because the food is visible near the mouth, which in turn increases **Q**11 scores and decreases **T**ime scores in the **O**n condition.
- **H**8 Increasing the downward FoV improves the ease of eating due to enhanced visibility of the food near the mouth, which in turn, results in an upward head pitch angle and reduces variance in the **O**n condition.

4.5 Participants

Participants were recruited through a campus mailing list and announcements. In accordance with the ethical review committee of the author's institution, informed consent was obtained from each of the participants after the study was fully explained to them. Each participant was paid the equivalent of 10 USD in local currency. Sixteen (eight males and eight females; mean age = 26.1 years; standard deviation = 8.89) participated in the study.

5 Results

The results are shown in Table 2 and Figs. 6, 7, 8, 9, 10, 11, and 12. In the figures, the red crosses represent the average values. Two-way ANOVA was conducted, considering the additional LCD units (ON vs. OFF) and the food appearance (before modulation (Sn) vs. after modulation (Fn)). Since the data did not follow a normal distribution, the aligned rank transform was used for hypothesis testing [39]. Significant differences are denoted by symbols (*** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$).

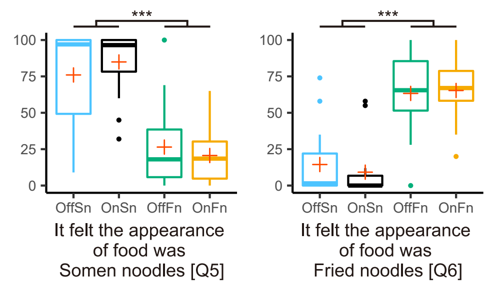

Figure 8: The perceived appearance of food results (Q5, Q6) [0 (strongly disagree) to 100 (strongly agree)].

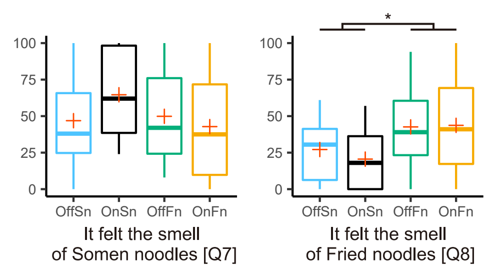

Figure 9: The perceived smell of food results (Q7, Q8) [0 (strongly disagree) to 100 (strongly agree)].

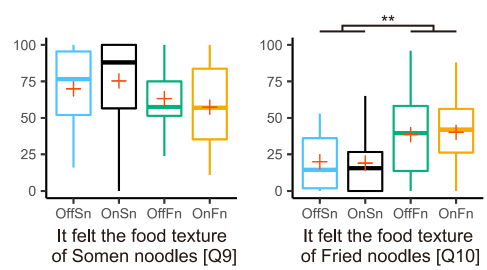

Figure 10: The perceived food texture results (Q9, Q10) [0 (strongly disagree) to 100 (strongly agree)].

Table 2 shows the Two-way ANOVA results and F-values, P-values, and Effect size η_p^2. Figure 6 shows the results of **Q**1 and **Q**2, which investigated the perceived taste of food. Two-way ANOVA revealed a significant main effect of the food appearance (Sn vs. Fn) on **Q**1 ($p < 0.01$) and **Q**2 ($p < 0.01$). No significant main effect of LCDs (ON vs. OFF) and interaction were found. Figure 7 shows the results of **Q**3 and **Q**4, which investigated the recognized type of food. Two-way ANOVA revealed a significant main effect

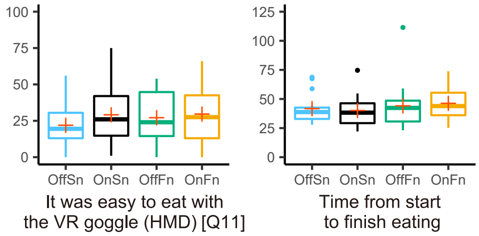

Figure 11: Left) The ease of eating results (Q11) [0 (strongly disagree) to 100 (strongly agree)]. Right) The time results while eating [s]

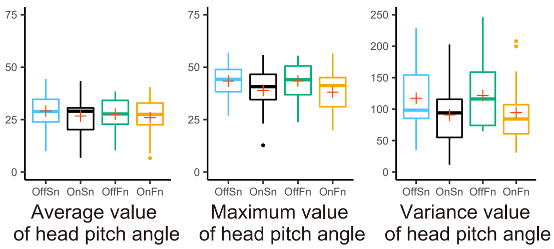

Figure 12: The head pitch angle while eating results [deg]. Left) Maximum value. Middle) Maximum value. Right) Variance value. The head pitch angle is 0 [deg] when facing front and +90 [deg] when fully facing down.

of the food appearance (Sn vs. Fn) on **Q**3 ($p < 0.01$) and **Q**4 ($p < 0.001$). No significant main effect of LCDs (ON vs. OFF) and interaction were found. Figure 8 shows the results of **Q**5 and **Q**6, which investigated the perceived appearance of food. Two-way ANOVA revealed a significant main effect of the food appearance (Sn vs. Fn) on **Q**5 ($p < 0.001$) and **Q**6 ($p < 0.001$). No significant main effect of LCDs (ON vs. OFF) and interaction were found. Figure 9 shows the results of **Q**7 and **Q**8, which investigated the perceived smell of food. Two-way ANOVA revealed a significant main effect of the food appearance (Sn vs. Fn) on **Q**8 ($p < 0.05$). No significant main effect of LCDs (ON vs. OFF) and interaction were found. Figure 10 shows the results of **Q**9 and **Q**10, which investigated the perceived food texture. Two-way ANOVA revealed a significant main effect of the food appearance (Sn vs. Fn) on **Q**10 ($p < 0.01$). No significant main effect of LCDs (ON vs. OFF) and interaction were found. Figure 11 shows the result of **Q**11 and the time results while eating, which investigated the ease of eating. Also, Figure 12 shows the result of head pitch angle, which investigated average, maximum, and variance. Two-way ANOVA did not find any significant differences in the **Q**11, time, and head pitch angle (Average, Maximum, Variance).

6 Discussion

The main findings are as follows.

1) The results from **Q**1 to **Q**4 indicate that changes in food appearance (Sn vs. Fn) have an impact on taste and food type perception. These findings align with previous studies [18, 19], suggesting that gustatory manipulation through visual modulation is achievable even when using HMDs with an increased downward FoV. However, there was no significant main effect of LCDs (ON vs. OFF), and no interaction was observed, which does not support **[**H1**]**. These results suggest that although an increased downward FoV improves visibility near the mouth, it may not necessarily affect the extent of gustatory manipulation through cross-modal effects. One reason could be that participants only occasionally saw the food until they brought it close to their mouths and consumed it. As a result, despite the increased visibility near the mouth, the amount of gustatory manipulation may not have changed because they did not observe the visually modulated food. Some participants felt positive about the expansion of the downward FoV, while others did not. Participants positively commented that the post-modulation food looked more like fried noodles when the downward FoV was expanded. Alternatively, some participants were concerned about the coarseness of the resulting image generated by the GAN, as the wider viewing angle highlighted the food. We plan to improve the gustation effect by improving the GAN quality.

An alternative explanation is that the cross-modal effect of visual modulation reached its maximum effectiveness in both conditions, resulting in no significant difference in the amount of gustatory manipulation. To further explore these hypotheses, it would be beneficial to incorporate eye trackers into HMDs with an increased downward FoV. This system would allow for measuring eye movements during eating and provide insights into the extent to which participants utilize the downward FoV. Although none of the participants reported any negative impact from the presence of the gap (several participants even compared it to the rim of a pair of glasses), it is worth considering that reducing or eliminating the gap could potentially enhance the overall user experience. In future work, we plan to address this issue by developing an improved system incorporating dedicated eyepieces, such as a fused lens [2].

Table 2: The results of the F and p values and effect sizes for the Two-way ANOVA. The food appearance conditions are abbreviated as “FA”. Items with a significant difference are in bold.

Question		Condition	F(1,60)	p	η_p^2
Taste	Somen noodles[Q1]	**FA**	**7.876**	**0.007**	**0.116**
		LCD	0.250	0.619	0.004
		FA:LCD	0.002	0.969	0.000
	Fried noodles[Q2]	**FA**	**10.852**	**0.002**	**0.153**
		LCD	0.277	0.601	0.005
		FA:LCD	0.025	0.875	0.000
Type	Somen noodles[Q3]	**FA**	**11.872**	**0.001**	**0.165**
		LCD	0.006	0.937	0.000
		FA:LCD	0.091	0.764	0.002
	Fried noodles[Q4]	**FA**	**20.713**	**0.000**	**0.257**
		LCD	0.001	0.979	0.000
		FA:LCD	0.108	0.743	0.002
Appearance	Somen noodles[Q5]	**FA**	**67.985**	**0.000**	**0.531**
		LCD	0.000	1.000	0.000
		FA:LCD	0.193	0.662	0.003
	Fried noodles[Q6]	**FA**	**74.628**	**0.000**	**0.554**
		LCD	0.272	0.604	0.005
		FA:LCD	0.130	0.720	0.002
Smell	Somen noodles[Q7]	FA	1.491	0.227	0.024
		LCD	0.305	0.583	0.005
		FA:LCD	2.277	0.137	0.037
	Fried noodles[Q8]	**FA**	**5.638**	**0.021**	**0.086**
		LCD	0.188	0.666	0.003
		FA:LCD	0.236	0.629	0.004
Texture	Somen noodles[Q9]	FA	3.734	0.058	0.059
		LCD	0.008	0.927	0.000
		FA:LCD	0.585	0.447	0.010
	Fried noodles[Q10]	**FA**	**9.675**	**0.003**	**0.139**
		LCD	0.021	0.886	0.000
		FA:LCD	0.263	0.610	0.004
Ease of eating[Q11]		FA	0.290	0.592	0.005
		LCD	0.712	0.402	0.012
		FA:LCD	0.223	0.638	0.004
Time		FA	0.957	0.332	0.016
		LCD	0.091	0.764	0.002
		FA:LCD	0.585	0.447	0.010
Head	Average	FA	0.050	0.824	0.001
		LCD	0.366	0.548	0.006
		FA:LCD	0.145	0.705	0.002
	Maximum	FA	0.021	0.886	0.000
		LCD	3.355	0.072	0.053
		FA:LCD	0.166	0.685	0.003
	Variance	FA	0.006	0.938	0.000
		LCD	3.792	0.056	0.059
		FA:LCD	0.050	0.824	0.001

2) The results of **Q**5 and **Q**6 indicate that the intended changes in food appearance (Sn and Fn) were successfully achieved. These results suggest that visual modulation is effectively applied in experiments using HMDs with an increased downward FoV. However, the absence of a significant main effect of LCDs (ON vs. OFF) and the lack of interaction do not support [H2]. These results suggest that the increased downward FoV does not necessarily enhance the perception of the visually presented food as more similar to the intended food. One potential explanation for this is a gap of approximately 10 [deg] between the front and bottom displays due to the 3D-printed housing, which prevents a seamless and borderless FoV.

3) The results of **Q**7 indicate that no significant main effect of food appearance (Sn vs. Fn) on perceived food smell was observed. However, the results of **Q**8 suggest that changes in the appearance of food (Sn vs. Fn) led participants to perceive the smell of fried noodles in the **F**n condition. These findings are interesting, as they differ from the cross-modal effect on gustation, which typically decreases the taste/type of the actual food and increases the taste/type of the presented food. These results partially support hypothesis [H3]. In summary, the cross-modal effect of visual modulation on olfaction does not diminish the smell of the actual food but enhances the smell of the visually presented food.

4) The results of **Q**9 indicate that no significant main effect of food appearance (Sn vs. Fn) on perceived food texture was observed. However, the results of **Q**10 suggest that changes in the appearance of food (Sn vs. Fn) led participants to perceive the food texture of fried noodles in the **F**n condition. These findings partially support hypothesis [H4]. In summary, similar to the findings on olfactory perception, the cross-modal effects of visual modulation on tactile perception enhanced the perceived food texture of visually presented food. Additionally, the results of **Q**9 show a marginal trend toward significance for food appearance ($p < 0.10$). This result suggests that changes in food texture may be more likely to occur than changes in food smell. Further studies should be conducted to determine how the cross-modal effects of vision on olfaction and tactile perception influence the perception of the actual smell and food texture.

5) The results from **Q**7 to **Q**10 show that there was no significant main effect of LCDs (ON vs. OFF) and interaction on perceived food smell and texture were found, which did not support [H5] and [H6]. These findings suggest that increasing the downward FoV to enhance mouth visibility does not seem to alter the magnitude of the cross-modal effect of visual modulation on tactile and olfaction, which aligns with the results obtained from **Q**1 to **Q**4.

6) The results of **Q**11 and **T**ime indicate that there were no significant main effects of LCDs (ON vs. OFF) and no interaction, which does not support hypothesis [H7]. Similarly, the results of **H**ead pitch angle show that there were no significant main effects of LCDs (ON vs. OFF) and no interaction, which do not support hypothesis [H8]. Contrary to expectations, these findings suggest that an increased downward visual field may not have an effect on the perceived ease of eating or the pitch angle of the head.

One possible explanation for the lack of change in participants' perception of the ease of eating (**Q**11) is that they may not have been looking at the food near their mouths while eating. Participants may have been able to visually confirm their ability to grasp the food with chopsticks on the front display and may not have required a downward FoV.

Regarding participants' comments on the ease of eating, we found a tendency for a mixture of participants who improved the ease of eating when the downward FoV was increased and those who felt it was difficult to eat even. The experimenter also observed that some participants ate with the bowl in their hands while others ate on the table. This difference in eating style may have increased or decreased the effect of the image display on the downward FoV.

We also believe that the lack of significant differences in the head pitch angle and time results may be due to the difficulty in setting the measurement interval correctly. Participants determined the end of the measurement period themselves because we could not visually confirm when they had finished eating. Furthermore, participants were free to move their head angle not only while eating but also when they were not directly looking at the food until they tasted it and completed the measurement. Future experiments should focus on measuring the head angle, specifically during the interval between bringing the food to their mouths and taking it into their mouths. In addition, although no significant differences were identified, the statistical results for the variance and maximum head movement in LCDs condition (ON vs. OFF) indicate a small effect size ($\eta_p^2 > 0.01$). These results suggest that the possibility that an enlarged downward FoV improves the ease of eating cannot be ruled out and requires further investigation.

7 Future work

The quantitative analysis could not show the expected effect and improvement of the downward FoV magnification in this study. However, most of the issues can be solved by reducing the lens gap of the HMD and improving the quality of the GAN, and further verification is possible by adding eye-tracking functions. We believe that the idea of visually modulating food under the user's eyes, which has been overlooked to date, is helpful in many gustatory researches

In this study, the low resolution of the GAN-generated images may have been made conspicuous by the wide viewing angle, which may have caused the taste to remain unchanged. Alternatively, the effects of food size manipulation without GANs are more likely to be felt near the mouth, which may improve the effectiveness of satiety manipulation [21]. It has also been shown that people tend to avoid high-calorie diets when using a slim avatar in a virtual environment [36]. Investigating the effect of avatar display on gustation in the downward FoV would be interesting. In addition, many studies have investigated whether the appearance of the virtual environment affects gustation, with some showing significant differences [2, 34] and others not [17, 24]. We expect that the idea of increasing the downward FoV will increase the realism and visibility near the mouth in virtual environments and increase the effectiveness of gustatory studies in the virtual environment. Many gustatory studies using VR/AR technology have the potential to improve people's health and quality of life caused by eating. In the future, people will use HMDs, just as they use smartphones while eating. When that future comes, we believe that it would be a waste to make people take off their HMDs because of the difficulty of eating, and we believe that it is necessary to communicate and continue the importance of research to reduce the difficulty of eating in gustatory studies using VR/AR technology.

8 Conclusion

In this study, we developed a VST-HMD with an increased downward FoV and examined its impact on cross-modal effects and the perceived ease of eating. However, the experimental results did not show significant differences in the amount of cross-modal effects or improvements in the ease of eating with the increased downward FoV. Nonetheless, the increased downward FoV allowed us to create a system to investigate the influence of food appearance near the mouth on perception. We found that visual modulation could change the perceived smell and food texture from the original food to the intended food. To further explore these findings, future studies are planned. One direction is to incorporate eye-tracking technology into the increased downward FoV HMD to examine participants' viewing patterns when bringing food to their mouths and understand how they perceive objects near their mouths. This direction will provide insights into the visual aspects of eating experiences. Overall, this study contributes to our understanding of the effects of visual modulation and the potential for cross-modal interactions in food perception.

ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI Grant Number JP20J21546 and JP22H00539.

REFERENCES

[1] J. J. Chai, C. O'Sullivan, A. A. Gowen, B. Rooney, and J. L. Xu. Augmented/mixed reality technologies for food: A review. *Trends in Food Science & Technology*, 124:182–194, 2022. doi: 10.1016/J.TIFS.2022.04.021

[2] Y. Chen, A. X. Huang, I. Faber, G. Makransky, and F. J. Perez-Cueto. Assessing the influence of visual-taste congruency on perceived sweetness and product liking in immersive vr. *Foods 2020, Vol. 9, Page 465*, 9:465, 4 2020. doi: 10.3390/FOODS9040465

[3] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of IEEE CVPR*, pp. 8789–8797, 2018.

[4] S. Damak, M. Rong, K. Yasumatsu, Z. Kokrashvili, V. Varadarajan, S. Zou, P. Jiang, Y. Ninomiya, and R. F. Margolskee. Detection of Sweet and Umami Taste in the Absence of Taste Receptor T1r3. *Science*, 301(5634):850–853, 2003. doi: 10.1126/science.1087155

[5] I. Endo, K. Takashima, M. Inoue, K. Fujita, K. Kiyokawa, and Y. Kitamura. A reconfigurable mobile head-mounted display supporting real world interactions. *Conference on Human Factors in Computing Systems - Proceedings*, 5 2021. doi: 10.1145/3411763.3451765

[6] R. P. Erickson. A Study of the Science of Taste: On the Origins and Influence of the Core Ideas. *Behavioral and Brain Sciences*, 31(1):59–75, 2008. doi: 10.1017/S0140525X08003348

[7] M. O. Ernst. A bayesian view on multimodal cue integration. *Human body perception from the inside out: Advances in visual cognition.*, pp. 105–131, 2006.

[8] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a

[9] Y. Ishiguro and J. Rekimoto. Peripheral vision annotation: Noninterference information presentation method for mobile augmented reality. In *Proceedings of the Augmented Human International Conference*, pp. 1–5. ACM Press, 2011. doi: 10.1145/1959826.1959834

[10] H. Iwata, H. Yano, T. Uemura, and T. Moriya. Food simulator: A haptic interface for biting. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 51–57, 2004. doi: 10.1109/VR.2004.1310055

[11] P. V. Joseph, D. R. Reed, and J. A. Mennella. Individual differences among children in sucrose detection thresholds: Relationship with age, gender, and bitter taste genotype. *Nursing Research*, 65:3, 2016. doi: 10.1097/NNR.0000000000000138

[12] N. Koizumi, H. Tanaka, Y. Uema, and M. Inami. Chewing jockey: augmented food texture by using sound based on the cross-modal effect. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, pp. 1–4, 2011. doi: 10.1145/2071423.2071449

[13] D. Korsgaard, T. Bjorner, J. R. Bruun-Pedersen, P. K. Sorensen, and F. J. Perez-Cueto. Eating together while being apart: A pilot study on the effects of mixed-reality conversations and virtual environments on older eaters' solitary meal experience and food intake. In *Proceedings of the IEEE VR*, pp. 365–370, 2020. doi: 10.1109/VRW50115.2020.00079

[14] J. J. W. Lin, H. B. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings of the Virtual Reality Annual International Symposium*, pp. 164–171, 2002. doi: 10.1109/vr.2002.996519

[15] R. W. Lindeman. A low-cost, low-latency approach to dynamic immersion in occlusive head-mounted displays. In *Proceedings of the Virtual Reality*, pp. 221–222. IEEE Computer Society, 2016. doi: 10.1109/VR.2016.7504733

[16] G. Morrot, F. Brochet, and D. Dubourdieu. The Color of Odors. *Brain and Language*, 79(2):309–320, 2001. doi: 10.1006/BRLN.2001.2493

[17] K. Nakano, D. Horita, N. Isoyama, H. Uchiyama, and K. Kiyokawa. Ukemochi: A video see-through food overlay system for eating experience in the metaverse. In *Proceedings of the CHI EA*, pp. 1–8, 2022. doi: 10.1145/3491101.3519779

[18] K. Nakano, D. Horita, N. Kawai, N. Isoyama, N. Sakata, K. Kiyokawa, K. Yanai, and T. Narumi. A Study on Persistence of GAN-Based Vision-Induced Gustatory Manipulation. *Electronics*, 10(10):1157, 2021. doi: 10.3390/ELECTRONICS10101157

[19] K. Nakano, D. Horita, N. Sakata, K. Kiyokawa, K. Yanai, and T. Narumi. DeepTaste: Augmented reality gustatory manipulation with GAN-based real-time food-to-food translation. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pp. 212–223. Institute of Electrical and Electronics Engineers Inc., 2019. doi: 10.1109/ISMAR.2019.000-1

[20] K. Nakano, N. Isoyama, D. Monteiro, N. Sakata, K. Kiyokawa, and T. Narumi. Head-Mounted Display with Increased Downward Field of View Improves Presence and Sense of Self-Location. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4204–4214, 2021. doi: 10.1109/TVCG.2021.3106513

[21] T. Narumi, Y. Ban, T. Kajinami, T. Tanikawa, and M. Hirose. Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 109–118, 2012. doi: 10.1145/2207676.2207693

[22] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose. Augmented reality flavors: Gustatory display based on Edible Marker and cross-modal interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 93–102, 2011. doi: 10.1145/1978942.1978957

[23] T. Narumi, M. Sato, T. Tanikawa, and M. Hirose. Evaluating cross-sensory perception of superimposing virtual color onto real drink: toward realization of pseudo-gustatory displays. In *Proceedings of the Augmented Human International Conference*, pp. 18–24, 2010. doi: 10.1145/1785455.1785473

[24] J. H. Oliver and J. H. Hollis. Virtual reality as a tool to study the influence of the eating environment on eating behavior: A feasibility study. *Foods 2021, Vol. 10, Page 89*, 10:89, 1 2021. doi: 10.3390/FOODS10010089

[25] K. Pennanen, J. Närväinen, S. Vanhatalo, R. Raisamo, and N. Sozer. Effect of virtual eating environment on consumers' evaluations of healthy and unhealthy snacks. *Food Quality and Preference*, 82:103871, 6 2020. doi: 10.1016/J.FOODQUAL.2020.103871

[26] J. Prescott. Multisensory processes in flavour perception and their influence on food choice. *Current Opinion in Food Science*, 3:47–52, 2015. doi: 10.1016/j.cofs.2015.02.007

[27] I. Rakkolainen, R. Raisamo, M. Turk, and T. Höllerer. Field-of-view extension for VR viewers. In *Proceedings of the International Academic Mindtrek Conference*, pp. 227–230. Association for Computing Machinery, Inc, 2017. doi: 10.1145/3131085.3131088

[28] N. Ranasinghe, T. N. T. Nguyen, Y. Liangkun, L.-Y. Lin, D. Tolley, and E. Y.-L. Do. Vocktail: A virtual cocktail for pairing digital taste, smell, and color sensations. In *Proceedings of the ACM international conference on Multimedia*, pp. 1139–1147, 2017. doi: 10.1145/3123266.3123440

[29] J. Ratcliff, A. Supikov, S. Alfaro, and R. Azuma. ThinVR: Heterogeneous microlens arrays for compact, 180 degree FOV VR near-eye displays. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1981–1990, 2020. doi: 10.1109/TVCG.2020.2973064

[30] S. D. Roper and N. Chaudhari. Taste buds: Cells, signals and synapses. *Nature Reviews Neuroscience*, 18(8):485–497, 2017. doi: 10.1038/nrn.2017.68

[31] P. Rozin. "Taste-smell confusions" and the duality of the olfactory sense. *Perception & Psychophysics*, 31(4):397–401, 1982. doi: 10.3758/BF03202667

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

[33] C. Spence, C. A. Levitan, M. U. Shankar, and M. Zampini. Does Food Color Influence Taste and Flavor Perception in Humans? *Chemosensory Perception*, 3(1):68–84, 2010. doi: 10.1007/s12078-010-9067-z

[34] A. Stelick, A. G. Penano, A. C. Riak, and R. Dando. Dynamic context

sensory testing–a proof of concept study bringing virtual reality to the sensory booth. *Journal of Food Science*, 83:2047–2051, 8 2018. doi: 10.1111/1750-3841.14275

[35] R. J. Stevenson, J. Prescott, and R. A. Boakes. Confusing tastes and smells: How odours can influence the perception of sweet and sour tastes. *Chemical Senses*, 24(6):627–635, 1999. doi: 10.1093/chemse/24.6.627

[36] R. Tambone, G. Poggio, M. Pyasik, D. Burin, O. D. Monte, S. Schintu, T. Ciorli, L. Lucà, M. V. Semino, F. Doricchi, and L. Pia. Changing your body changes your eating attitudes: embodiment of a slim virtual avatar induces avoidance of high-calorie food. *Heliyon*, 7, 2021. doi: 10.1016/j.heliyon.2021.e07515

[37] K. T. Tran, S. Jung, S. Hoermann, and R. W. Lindeman. MDI: A Multi-channel Dynamic Immersion Headset for Seamless Switching between Virtual and Real World Activities. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pp. 350–358. IEEE, 2019. doi: 10.1109/VR.2019.8798240

[38] J. Ueda and K. Okajima. AR food changer using deep learning and cross-modal effects. In *Proceedings of IEEE AIVR*, pp. 110–117, 2019. doi: 10.1109/AIVR46125.2019.00025

[39] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 143–146. ACM Press, 2011. doi: 10.1145/1978942.1978963

[40] R. Xiao and H. Benko. Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1221–1232. ACM Press, 2016. doi: 10.1145/2858036.2858212

[41] W. Yamada and H. Manabe. Expanding the Field-of-View of Head-Mounted Displays with Peripheral Blurred Images. In *Proceedings of the Annual Symposium on User Interface Software and Technology*, pp. 141–142. ACM Press, 2016. doi: 10.1145/2984751.2985735

[42] M. Zampini and C. Spence. The Role of Auditory Cues in Modulating The Perceived Crispness and Staleness of Potato Chips. *Journal of Sensory Studies*, 19(5):347–363, 2004. doi: 10.1111/j.1745-459x.2004.080403.x