2026 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEP. 28– OCT. 1, 2026, ATLANTA, USA

SonoBite: ULTRASOUND SENSING FOR SMARTPHONE-BASED FOOD RECOGNITION

LaxmiPrasanna Ravikanti[†], *Srinivasan Murali*[‡], *Jianshu Liu*[§], *Shuaikang Hou*[†], *Huadi Zhu*[†*]

[†] Georgia State University, [‡] Texas A&M University-San Antonio, [§] San Diego State University

ABSTRACT

Continuous dietary monitoring is important for preventive healthcare and personalized nutrition management, yet existing food recognition systems often rely on wearable devices or camera-based sensing that limit usability and raise privacy concerns. In this paper, we present *SonoBite*, a smartphone-based food recognition system using active ultrasound acoustic sensing and deep learning. *SonoBite* transforms commodity smartphones into active sensing platforms by emitting inaudible ultrasound chirps and analyzing chewing-induced signal reflections during food consumption. The reflected signals are converted into STFT spectrograms and processed using a deep neural network trained with domain generalization for robust food recognition. We evaluate *SonoBite* through extensive experiments involving 20 participants across diverse real-world environments. Results show that *SonoBite* improves the state-of-the-art baseline by 7% accuracy while maintaining robustness across heterogeneous users, devices, sensing distances, and environmental conditions. Our study demonstrates the feasibility of privacy-preserving dietary sensing using ubiquitous mobile devices.

***Index Terms*—** Mobile health, acoustic sensing, signal processing, deep learning

1. INTRODUCTION

Dietary habits are closely associated with human health and quality of life. Poor eating behaviors contribute to numerous chronic conditions, including obesity, diabetes, and cardiovascular diseases [1]. As a result, continuous dietary monitoring has become increasingly important for preventive healthcare, nutrition assessment, and health management. However, developing a practical and scalable food recognition system for daily deployment remains challenging.

Recent advances in mobile sensing and machine learning have enabled new opportunities for automatic dietary monitoring. Existing systems have explored wearable acoustic sensors, inertial sensing, and camera-based approaches for detecting eating activities and recognizing food types [2, 3, 4, 5]. Despite promising progress, most prior solutions still suffer from several limitations. Wearable-based approaches require users to continuously wear dedicated devices such as earables or body-mounted sensors, limiting long-term usability and accessibility. Camera-based methods introduce substantial privacy concerns and high energy consumption due to continuous visual capture. Meanwhile, smartphones, despite being ubiquitous and sensor-rich platforms, remain largely underexplored for fine-grained food recognition.

In this paper, we investigate whether commodity smartphones alone can support robust food recognition through ultrasound sensing of chewing behaviors. Chewing produces subtle mouth and jaw movements that vary according to food texture, hardness, and consumption dynamics [6]. These movements modulate reflected acoustic signals and generate distinctive temporal-frequency patterns. Motivated by this observation, we propose *SonoBite*, a smartphone-based food recognition system using active ultrasound acoustic sensing and deep spectrogram learning.

SonoBite transforms smartphones into active sensing platforms by emitting inaudible ultrasound chirp signals and analyzing their reflections during food consumption. Compared with passive acoustic sensing, active ultrasound sensing provides stronger robustness against environmental interference and preserves user privacy by avoiding speech or image capture. The reflected signals are converted into STFT spectrograms, which are then processed using a deep neural network to learn discriminative chewing representations for food recognition.

Designing such a system introduces several technical challenges. First, smartphone-based ultrasound sensing is susceptible to interference from direct-path leakage, multipath reflections, and environmental noise. Second, chewing patterns exhibit substantial variability across users, food categories, smartphone hardware, and sensing environments. Third, mobile deployment requires lightweight and efficient learning pipelines that can execute continuously on resource-constrained devices.

To address these challenges, *SonoBite* incorporates an end-to-end ultrasound sensing and learning framework consisting of signal preprocessing, STFT spectrogram generation, domain-aware deep learning, and efficient mobile inference. We further investigate robustness across heterogeneous sensing domains including different smartphone

* Corresponding author

979-8-3195-0884-3/26/$31.00 ©2026 IEEE

models, placements, rooms, and ambient noise conditions.

We evaluate *SonoBite* through extensive experiments involving 20 participants across diverse real-world environments. Results demonstrate that *SonoBite* improves state-of-the-art solution by 7% accuracy and achieves robust food recognition performance while maintaining low runtime overhead on smartphones. Our study highlights the potential of combining deep learning and ultrasound signal processing for dietary monitoring on ubiquitous mobile devices.

In summary, this work makes the following contributions:

- We present *SonoBite*, a smartphone-based food recognition system using active ultrasound sensing of chewing behaviors.
- We design an end-to-end spectrogram learning framework that combines ultrasound signal preprocessing, STFT representation learning, and domain-aware deep neural networks for robust food recognition.
- We conduct extensive evaluations, which demonstrate the feasibility and robustness of ultrasound-based dietary sensing on commodity smartphones.

2. RELATED WORK

Chewing behaviors encode rich information about food texture, hardness, and consumption patterns through coordinated mouth and jaw movements [7]. Prior healthcare studies have shown strong correlations between chewing behaviors and dietary quality, nutritional intake, and metabolic health [1, 8]. Motivated by these findings, recent sensing systems have explored chewing as a behavioral signal for dietary monitoring.

Existing approaches mainly rely on wearable sensing modalities, including acoustic [9, 2], inertial [10, 4], piezoelectric [11, 12], and hybrid sensing systems [13, 14]. Early systems primarily focused on coarse-grained eating activity detection using custom hardware or body-worn sensors. More recent works improved practicality by leveraging commodity wearable devices.

In parallel, camera-based dietary monitoring systems have also been widely explored [15, 16, 5]. These approaches rely on RGB or infrared imaging to recognize eating episodes. While effective in controlled environments, continuous visual sensing introduces substantial privacy concerns and energy overhead, limiting practicality for daily deployment.

Compared with wearable systems, smartphones remain surprisingly underexplored for dietary sensing despite their ubiquity and rich sensing capabilities. Existing smartphone approaches primarily rely on passive sensing or self-reported user input [17, 18], which limits sensing fidelity and automation capability.

Among existing systems, *BiteSense* [4] is the closest and most relevant prior work to ours. To the best of our knowledge, it is also the only recent state-of-the-art system that supports fine-grained food recognition using inertial sensing. We therefore use *BiteSense* as our primary baseline in evaluation. In contrast to wearable-based approaches, our work investigates whether fine-grained food recognition can be achieved directly using commodity smartphones through active ultrasound sensing without requiring dedicated wearable devices.

3. SYSTEM ARCHITECTURE

3.1. Overview

We present *SonoBite*, a smartphone-based food recognition system that leverages ultrasound acoustic sensing and deep learning to recognize food consumption behaviors from chewing patterns. Unlike prior approaches that rely on wearable sensors or camera-based monitoring, *SonoBite* operates entirely on commodity smartphones and preserves user privacy by sensing only ultrasound reflections rather than visual or speech content.

Figure 1 illustrates the overall workflow of *SonoBite*. During food consumption, the smartphone continuously emits inaudible ultrasound chirp signals through its built-in speaker while simultaneously recording reflected signals using the microphone. Jaw and mouth movements during chewing perturb the reflected acoustic waveforms, producing distinct temporal-frequency patterns associated with different food types. The captured signals are preprocessed to suppress environmental interference and direct-path leakage, then transformed into Short-Time Fourier Transform (STFT) spectrograms. Finally, a deep neural network performs food recognition based on these spectrogram representations.

The design of *SonoBite* is motivated by two key observations. First, different foods induce distinct chewing dynamics due to variations in texture, hardness, and chewing rhythm. Second, ultrasound sensing provides high temporal sensitivity to subtle mouth motions while avoiding many privacy concerns associated with camera or speech sensing. Building on these observations, *SonoBite* formulates food recognition as a signal representation learning problem over spectrograms.

3.2. Active Ultrasound Sensing

Signal design. To capture fine-grained chewing dynamics, *SonoBite* adopts an active acoustic sensing approach. Compared with passive acoustic sensing, it provides stronger robustness against ambient sound interference and captures subtle motion variations at high temporal resolution. The smartphone periodically emits linear chirp signals, defined as:

$$S(t) = A \cdot \cos\left(\frac{\pi(f_h - f_l)}{T}t^2 + 2\pi f_l t\right) \quad (1)$$

where A denotes the signal amplitude, T is the chirp duration, and f_h and f_l represent the upper and lower frequencies, respectively. Following prior ultrasound sensing systems

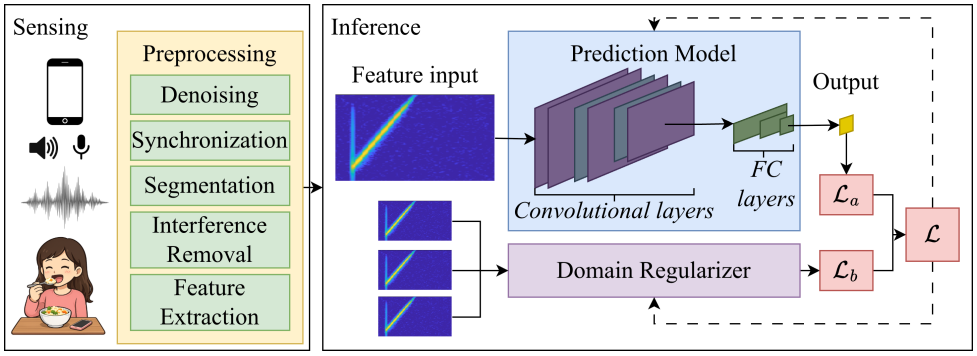


Fig. 1: System architecture.

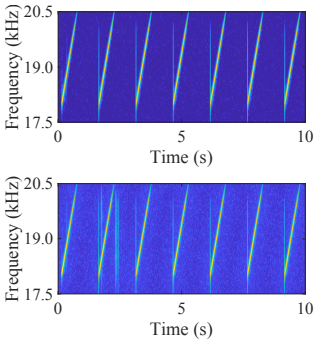


Fig. 2: Example signals: water (top) vs tortilla (bottom).

[19, 20], we use the 18–20 kHz frequency range to balance sensing fidelity and inaudibility for daily use.

Signal preprocessing. Raw reflected signals contain interference from environmental noise, direct speaker-to-microphone leakage, and surrounding reflections. To improve signal quality, *SonoBite* first applies a Butterworth band-pass filter to retain only the 17.5–20.5 kHz frequency components. The filtered signals are then synchronized with emitted chirps and segmented into fixed-length sensing windows of 1.5 seconds, each segment covering a complete chirp.

A primary challenge is the impact of interferences, which come from two major sources, the user's body (S_u^i) and other ambient reflections (S_a^i), for the propagation path i. We first measure typical phone-user distances, which remain stable across users/scenes. We then apply a multi-path model for the received chirp after propagation:

$$r(t) = \sum_{i \in \mathcal{N}} H(S_u^i(t-\tau) + S_u^i(t-\tau)) \tag{2}$$

where H denotes the set of decay coefficients of all propagation paths with a path delay τ in the observed channel impulse response. With S_u^i known *a-priori*, we estimate:

$$H^* = \arg\min_H : \sum_t \left[r(t) - \sum_{i \in \mathcal{N}} \left(H \cdot S_a^i \right) \right]^2 . \tag{3}$$

3.3. Food Recognition

Feature extraction. After preprocessing, *SonoBite* converts each signal segment into a time-frequency representation using the Short-Time Fourier Transform (STFT). Figure 2 shows example spectrograms generated from different food consumption activities. Distinct chewing behaviors produce different spectral textures and temporal patterns in the reflected ultrasound signals. The entire 2D spectrogram is fed to the inference model as feature input.

Deep learning framework. We formulate food recognition as a multi-class classification problem over STFT spectrograms. To learn discriminative chewing representations, *SonoBite* employs a lightweight convolutional neural network (CNN) consisting of stacked convolutional and max-pooling layers followed by fully connected classification layers.

One key challenge in real-world deployment is environmental and hardware variability. Different smartphone models, device placements, and room layouts introduce substantial domain shifts in captured ultrasound signals, which can degrade model generalization. To address this issue, *SonoBite* incorporates domain-aware representation learning during training. Specifically, training data are collected across diverse devices, placements, and environments. We introduce a domain regularization objective that encourages feature consistency across different sensing domains:

$$\mathcal{L}_b = \sum_{i \neq j} (D_{KL}(f_i, f_j) + D_{KL}(f_j, f_i)) \tag{4}$$

where f_i and f_j denote feature embeddings from different domains and $D_{KL}(\cdot)$ represents the Kullback-Leibler divergence. The final training objective combines the classification loss and domain regularization term:

$$\mathcal{L} = \mathcal{L}_a + \lambda \cdot \mathcal{L}_b \tag{5}$$

where λ controls the strength of domain regularization.

4. EVALUATION

4.1. Experiment Setup

We evaluate *SonoBite* through a large-scale food consumption study involving 20 participants under diverse sensing conditions. Participants were recruited from three universities and completed supervised data collection sessions using commodity Android smartphones. During each session, participants consumed multiple categories of food and beverages while placing the smartphone on a nearby table surface. To improve data diversity and reduce ordering bias, food assignments and consumption sequences were randomized across users. Each recording session lasted 30 seconds and contained continuous natural eating behaviors. The study protocol was IRB-approved.

Table 1: Food types and properties.

Index	Type	State	Texture	Nutrition
1	Tortilla	Solid	Soft	Carb
2	Fruit	Solid	Soft	Carb
3	Chicken	Solid	Soft	Protein
4	Cracker	Solid	Crunchy	Carb+Protein
5	Carrot	Solid	Crunchy	Carb
6	Chocolate	Solid	Crunchy	Fatty
7	Yogurt	Semi-liquid	Soft	Protein
8	Noodles	Semi-liquid	Soft	Carb
9	Water	Liquid	N/A	N/A
10	Coke	Liquid	N/A	Fatty

Overall, the dataset contains 34,360 sensing segments, each lasting 1.5 seconds, collected across heterogeneous users, environments, and device configurations. The dataset is train-test split by segments at 80:20. Table 1 summarizes the evaluated food categories. The dataset intentionally includes foods with different and mixed physical properties to evaluate the discriminative capability of ultrasound sensing.

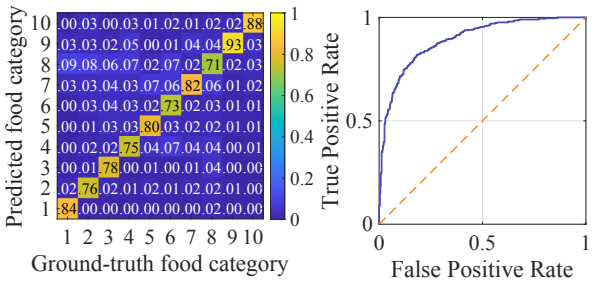
4.2. Overall Performance

Recognition performance. We first evaluate the overall food recognition capability of *SonoBite*. Figure 3a presents the confusion matrix across all food categories. The system achieves strong classification performance for most food types. Foods with distinctive physical properties, particularly liquids such as water, achieve the highest recognition accuracy due to their unique swallowing and chewing patterns. In contrast, semi-liquid foods such as noodles exhibit comparatively lower accuracy because they combine heterogeneous chewing and swallowing dynamics. A similar result was reported in [4]. Figure 3b shows the ROC performance of *SonoBite*, which achieves an AUC of 0.89. Overall, the results demonstrate that ultrasound signals capture sufficiently discriminative temporal-frequency patterns for fine-grained food recognition.

Baseline comparison. Compared with the state-of-the-art solution *BiteSense* [4], *SonoBite* demonstrates better accuracy (0.81 vs 0.76)[1], while eliminating the need for body-worn sensors or camera-based monitoring. This demonstrates the effectiveness of combining ultrasound sensing and deep spectrogram learning for privacy-preserving dietary analysis.

Ablation study. To assess the contribution of domain regularization, we conduct an ablation study by eliminating this component and remeasuring the system performance. We observe an accuracy drop from 0.81 to 0.69, caused by the increased impact of devices, placements, and environments.

[1]We adopt the accuracy reported in [4] as its code is not publicly available.



(a) Confusion matrix. (b) ROC curve.

Fig. 3: Overall performance of *SonoBite*.

This suggests the importance of domain regularization.

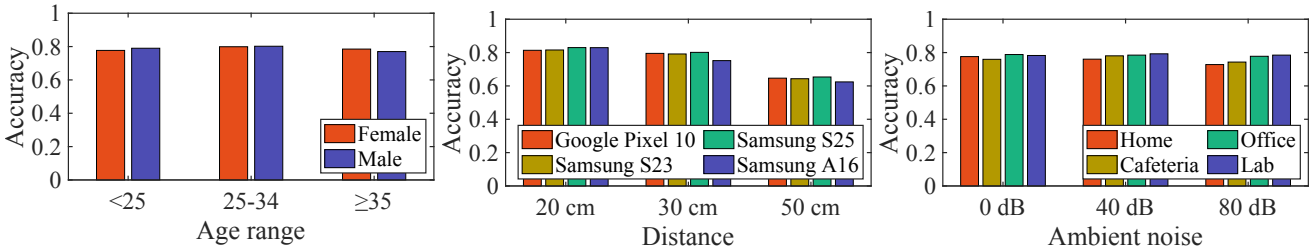
4.3. Generalization Across Sensing Domains

A major challenge for smartphone-based acoustic sensing systems is domain variability introduced by users, devices, and environments. We therefore evaluate the robustness of *SonoBite* across heterogeneous sensing domains.

User groups. Figure 4a shows the recognition performance across different user groups. Despite natural variations in chewing habits and eating behaviors, *SonoBite* maintains relatively stable performance across demographic groups. This suggests that the learned spectrogram representations capture food-dependent characteristics rather than overfitting to individual users.

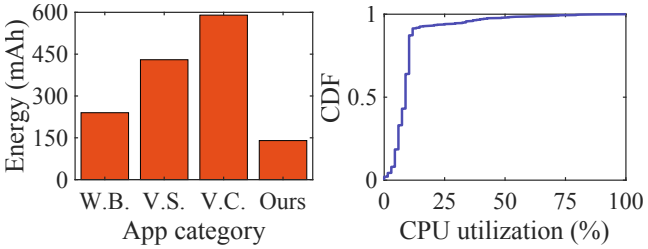
Smartphone models and distances. We further evaluate *SonoBite* across multiple smartphone models with different speaker–microphone layouts and sensing capabilities. As shown in Figure 4b, newer devices generally achieve slightly higher accuracy due to improved acoustic hardware, although all tested smartphones maintain satisfactory performance. We also vary the sensing distance between the user and smartphone to emulate realistic placement conditions. Recognition accuracy gradually decreases at longer distances because of ultrasound attenuation and reduced signal-to-noise ratio. Nevertheless, *SonoBite* remains robust under practical table-level deployment distances.

Room configurations and noise levels. To examine environmental generalization, we conduct experiments in multiple indoor environments including home, cafeteria, office, and lab. We additionally inject ambient white noise at different intensity levels. Figure 4c shows that *SonoBite* experiences only minor degradation across environmental conditions. This robustness mainly stems from two factors: (1) ultrasound sensing operates in high-frequency bands that are less affected by common environmental sounds, and (2) deep learning with domain generalization improves feature consistency across heterogeneous sensing environments.



(a) User groups (b) Device types and distances (c) Room types and noise levels

Fig. 4: Consistency test results.



(a) Energy consumption (b) CPU utilization

Fig. 5: System runtime overhead.

4.4. System Overhead

Energy consumption. Continuous sensing systems must operate under limited battery budgets. Figure 5a shows the average energy consumption of *SonoBite* during continuous operation. *SonoBite* consumes only 140 mAh on average over 30 min, translating to approximately 3.5% battery drop. This is much smaller than common applications such as web browsing (≈240 mAh), video streaming (≈430 mAh), and video calling (≈590 mAh).

Computational overhead. We further monitor CPU utilization during real-time sensing and inference over a 20-minute experimental session using the Android Debug Bridge (ADB). As shown in Figure 5b, CPU utilization remains below 10% for >90% of the runtime with an average of 8.2%, indicating low computational overhead and practical feasibility for continuous deployment.

5. LIMITATIONS AND FUTURE WORK

Despite encouraging results, several limitations remain. First, the current study evaluates a limited set of food categories under controlled consumption sessions. Real-world dietary behaviors may involve more complex eating patterns, mixed foods, and unconstrained environmental dynamics. Second, sensing performance can be affected by different smartphone pointing angles relative to the user. Variations in orientation change the ultrasound propagation paths and introduce additional multipath reflections, which may distort chewing-related signal patterns and reduce recognition robustness.

Future work will explore larger-scale longitudinal deployments and more diverse real-world dietary scenarios. We also plan to investigate adaptive beamforming, angle-aware signal calibration, and multipath-resilient learning techniques to improve robustness against orientation and reflection variations.

6. CONCLUSION

This paper presented *SonoBite*, a smartphone-based food recognition system using ultrasound acoustic sensing and deep learning. *SonoBite* transforms commodity smartphones into privacy-preserving sensing platforms by emitting inaudible ultrasound chirps and analyzing chewing-induced signal reflections through STFT spectrogram learning. To address the challenges of interferences and heterogeneous sensing environments, we designed a mobile sensing pipeline and a deep learning framework assisted with domain generalization. Extensive experiments involving 20 participants demonstrate that *SonoBite* improves state-of-the-art by 7% and achieves robust food recognition performance across diverse users, devices, and environmental conditions while maintaining low runtime overhead on smartphones. Our results demonstrate the potential of combining ultrasound sensing and deep spectrogram learning for scalable mobile dietary monitoring.

7. REFERENCES

[1] Dariush Mozaffarian, "Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review," *Circulation*, vol. 133, no. 2, pp. 187–225, 2016.

[2] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al., "Auracle: Detecting eating episodes with an ear-mounted sensor," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–27, 2018.

[3] Tamisa Ketmalasiri, Yu Yvonne Wu, Kayla-Jade Butkow, Cecilia Mascolo, and Yang Liu, "Imchew: Chewing analysis using earphone inertial measurement units," in *Proceedings of the Workshop on Body-Centric Computing Systems*, 2024, pp. 29–34.

[4] Garvit Chugh, Indrajeet Ghosh, Sandip Chakraborty, and Suchetana Chakraborty, "Bitesense: Earable-based inertial sensing for eating behaviour assessment," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2025, pp. 110–120.

[5] Soroush Shahi, Sougata Sen, Mahdi Pedram, Rawan Alharbi, Yang Gao, Aggelos K Katsaggelos, Josiah Hester, and Nabil Alshurafa, "Detecting eating, and social presence with all day wearable rgb-t," in *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, 2023, pp. 68–79.

[6] Kartika Indah Sari and Anggun Rafisa, "Chewing and swallowing patterns for different food textures in healthy subjects," *International Journal of Dentistry*, vol. 2023, no. 1, pp. 6709350, 2023.

[7] Ronald Dubner, *The neural basis of oral and facial function*, Springer Science & Business Media, 2013.

[8] Shuri Fushida, Takayuki Kosaka, Michikazu Nakai, Momoyo Kida, Takashi Nokubi, Yoshihiro Kokubo, Makoto Watanabe, Yoshihiro Miyamoto, Takahiro Ono, and Kazunori Ikebe, "Lower masticatory performance is a risk for the development of the metabolic syndrome: the suita study," *Frontiers in Cardiovascular Medicine*, vol. 8, pp. 752667, 2021.

[9] Koji Yatani and Khai N Truong, "Bodyscope: a wearable acoustic sensor for activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 341–350.

[10] Oliver Amft and Gerhard Troster, "On-body sensing solutions for automatic dietary monitoring," *IEEE pervasive computing*, vol. 8, no. 2, pp. 62–70, 2009.

[11] Tauhidur Rahman, Alexander Travis Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury, "Bodybeat: a mobile system for sensing non-speech body sounds," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, 2014, vol. 14, pp. 2–594.

[12] Jaemin Shin, Seungjoo Lee, Taesik Gong, Hyungjun Yoon, Hyunchul Roh, Andrea Bianchi, and Sung-Ju Lee, "Mydj: Sensing food intakes with an attachable on your eyeglass frame," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–17.

[13] Jindong Liu, Edward Johns, Louis Atallah, Claire Pettitt, Benny Lo, Gary Frost, and Guang-Zhong Yang, "An intelligent food-intake monitoring system using wearable sensors," in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*. IEEE, 2012, pp. 154–160.

[14] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg, "Recognizing eating from body-worn sensors: Combining free-living and laboratory data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–20, 2017.

[15] Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen, "Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype," in *Proceedings of the 4th Workshop on Embedded networked sensors*, 2007, pp. 13–17.

[16] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel, "Fitbyte: Automatic diet monitoring in unconstrained situations using multimodal sensing on eyeglasses," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[17] Edmund Seto, Jenna Hua, Lemuel Wu, Victor Shia, Sue Eom, May Wang, and Yan Li, "Models of individual dietary behavior based on smartphone data: the influence of routine, physical activity, emotion, and food environment," *PLoS One*, vol. 11, no. 4, pp. e0153085, 2016.

[18] Lakmal Meegahapola, Salvador Ruiz-Correa, Viridiana del Carmen Robledo-Valero, Emilio Ernesto Hernandez-Huerfano, Leonardo Alvarez-Rivera, Ronald Chenu-Abente, and Daniel Gatica-Perez, "One more bite? inferring food consumption level of college students using smartphone sensing and self-reports," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–28, 2021.

[19] Wenguang Mao, Mei Wang, and Lili Qiu, "Aim: Acoustic imaging on a mobile," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 468–481.

[20] Wenqiang Jin, Srinivasan Murali, Youngtak Cho, Huadi Zhu, Tianhao Li, Rachael Thompson Panik, Anika Rimu, Shuchisnigdha Deb, Kari E Watkins, Xu Yuan, et al., "Cycleguard: A smartphone-based assistive tool for cyclist safety using acoustic ranging.," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 4, pp. 163–1, 2021.