Capturing Chewing and Swallowing with Earables: A Multimodal Dataset Across Contexts

Jun Fang*
Department of Computer Science and Technology, Tsinghua University
China
fangy23@mails.tsinghua.edu.cn

Ka I Chan*
School of Information, University of Michigan
Ann Arbor, United States
chankai@umich.edu

Xiyuxing Zhang*
Department of Computer Science and Technology, Tsinghua University
China
zxyx22@mails.tsinghua.edu.cn

Yuntao Wang†
Key Laboratory of Pervasive Computing, Ministry of Education, Department of Computer Science and Technology, Tsinghua University
China
yuntaowang@tsinghua.edu.cn

Zihang Zhan
Zhixin Zhao
zhanzh22@mails.tsinghua.edu.cn
zhaozhix22@mails.tsinghua.edu.cn
Tsinghua University
China

Yuanchun Shi
Key Laboratory of Pervasive Computing, Ministry of Education, Department of Computer Science and Technology, Tsinghua University
China
Intelligent Computing and Application Laboratory of Qinghai Province, Qinghai University
China
shiyc@tsinghua.edu.cn

Abstract

Fine-grained ingestive events such as chewing and swallowing provide a direct window into how a meal unfolds and can enable actionable feedback on eating pace. However, collecting temporally precise labels in natural meals remains difficult, and many existing datasets rely on camera-centric ground truth or controlled conditions. We present a multimodal dataset centered on commodity earables that capture in-ear bone-conducted audio, augmented with laryngophone audio, bilateral wrist IMUs, and two video views (egocentric and frontal). The frontal video serves as the primary ground-truth reference for annotation, while synchronized laryngophone audio and egocentric video provide complementary cues when events are ambiguous. We collect 10.07 hours of recordings from 18 participants across two contexts, a quiet laboratory room and a free-living campus cafeteria, and provide event-level annotations totaling 45,582 chewing events and 2,150 swallows. We additionally report concise baselines to illustrate current capability and open challenges, including a clear modality trade-off between chewing and swallowing recognition and sensitivity of swallowing detection to temporal segmentation. We expect the dataset to support benchmarking and future work on robust ingestive sensing from lab to free living, modality trade-offs, and multimodal fusion.

*These authors contributed equally to this research.
†Corresponding authors.

CHI '26, Barcelona, Spain

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing**.

Keywords

wearable sensing; chewing; swallowing; earables; bone-conduction audio; free-living; multimodal dataset

ACM Reference Format:
Jun Fang, Ka I Chan, Xiyuxing Zhang, Yuntao Wang, Zihang Zhan, Zhixin Zhao, and Yuanchun Shi. 2026. Capturing Chewing and Swallowing with Earables: A Multimodal Dataset Across Contexts. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing System (CHI '26).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Eating is a foundational daily activity whose effects extend to health, quality of life, and long-term well-being [3, 10]. Beyond what people eat, how they eat, such as chewing counts and eating pace, also shapes satiety signals, meal enjoyment, and the ability to self-regulate intake [2, 15]. HCI research has started to build datasets and models for eating behavior to help people understand and improve dietary habits, yet many studies remain at a coarse granularity. At the coarse level, systems identify eating episodes from daily activities such as walking, speaking, and working, which supports longitudinal reflection on everyday intake patterns [16, 20]. In contrast, fine-grained modeling focuses on micro-events within a meal, including chewing cycles, swallowing actions, drinking behaviors, and hand-to-mouth gestures, to explain how eating unfolds over time. Fine-grained sensing is critical for actionable feedback, such as detecting rapid eating or irregular pacing, but it requires temporally precise and richly annotated data. Models developed in controlled environments often perform poorly in free-living meals because of ambient noise, conversational speech, body motion, and

differences in food texture. These practical needs motivate datasets that include both laboratory and free-living recordings and provide multimodal reference signals that support reliable annotation.

However, existing fine-grained datasets often rely on cameras as the primary reference [12, 14] and focus on distinguishing bodily behaviors during eating. Continuous deployment in natural meals remains difficult because of privacy constraints, occlusion, and social acceptability. In addition, many datasets depend on research-grade devices and controlled data collection, which is still far from everyday, wearable, and practical ubiquitous computing. Earbuds are increasingly common, and their stable placement captures body-conducted vibrations induced by jaw motion [17]. Ear-worn audio modules can also capture chewing and swallowing sounds, which motivates a growing body of work that uses eating-related audio [11]. This sensing pathway is promising for daily eating monitoring, yet publicly described ingestive datasets that cover both laboratory and free-living contexts and combine ear-worn signals with multimodal references and high-fidelity ground truth remain limited.

In this paper, we present a multimodal dataset for fine-grained eating behavior analysis based on commercial earbuds with complementary sensing. The dataset includes synchronized recordings of in-ear bone-conduction audio from commodity earbuds, laryngophone audio as a high-fidelity reference, IMU signals from both wrists during eating, and two video views, namely egocentric video and frontal (eye-level) video, where the frontal view serves as the primary ground-truth reference for annotation. We collect data from 18 healthy adults in two contexts, a quiet laboratory room and a participant-selected campus cafeteria during regular meal times, where participants freely choose their food and eat at their natural pace. Across both contexts, the dataset contains 10.07 hours of recordings (5.65 hours in-lab and 4.42 hours in free living). Using the frontal video as the primary reference and consulting the synchronized egocentric video and laryngophone audio when needed, we align the streams and manually annotate chewing and swallowing events in the earbud signal, yielding a total of 45,582 chewing events and 2,150 swallows. We expect this dataset to support downstream analysis and benchmarking on robustness from lab to free living, modality trade-offs among earable sensing, throat audio, and fusion, and the remaining challenges of fine-grained ingestive sensing in everyday meals.

This work contributes an aligned and annotated multimodal dataset centered on ear-worn sensing of chewing and swallowing across laboratory and free-living settings, along with some preliminary findings across multiple modeling approaches and sensor configurations that reflects current performance and open limitations.

2 Related Works and Background

Public datasets have become a key driver for modeling eating-related activities, especially for fine-grained intake events. OREBA provides synchronized video and wrist-worn inertial measurement unit signals from both hands, along with thousands of annotated intake gestures that support gesture-level recognition [14]. FIC dataset focuses on within-meal eating behavior using wrist inertial sensing [6], and it is later extended to FreeFIC to better reflect eating in natural environments [4, 5, 7]. EatSense further segments eating behavior into multiple sub-actions based on skeleton-processed video [12]. While these datasets are valuable, they are often camera-centric or wrist-centric and primarily target hand-to-mouth actions and related cues, which may not fully capture fine-grained oropharyngeal events such as chewing and swallowing that are critical for understanding how a meal unfolds.

Beyond vision-based approaches, neck and throat sensing, including laryngophones and accelerometer-based cervical auscultation, has a long history of detecting swallow-related vibrations and sounds and is widely regarded as a promising non-invasive direction for swallow detection [18, 19]. Recent work also explores sensing closer to the jaw and throat region through everyday wearables such as glasses and earphones to support fine-grained eating monitoring [1, 17]. Earables can capture chewing-related characteristics through acoustics while maintaining social acceptability for daily use [13]. However, datasets that combine commodity ear-worn sensing with throat or neck references and cover both controlled and free-living meals remain limited.

Moving from laboratory studies to real-world deployment is widely recognized as necessary but challenging. Evidence from in-the-wild physiological datasets shows that models trained in controlled settings often degrade in everyday contexts, and collecting labeled longitudinal or free-living data requires strong participant adherence [11]. Free-living sensing also introduces noise and missing data due to device connectivity issues, improper wear, and non-wear periods [8]. These challenges motivate datasets that prioritize cross-context multimodal synchronization and event-level annotation. For fine-grained eating behavior in particular, existing earable datasets often rely on air-conducted audio collected in controlled settings [9, 11], where background noise is less severe than in real meals. In this work, we instead use commodity earables with bone-conduction sensors together with a laryngophone, the wrist-worn IMU and reference video to collect and annotate fine-grained ingestive events across controlled and free-living contexts, providing a complementary resource to existing gesture- and camera-centric datasets.

3 Dataset

Motivated by these gaps, we present a multimodal dataset and describe its data collection process and event-level annotation pipeline in this section.

3.1 Data Collection

Our dataset focuses on fine-grained ingestive events, specifically chewing and swallowing. We build a lightweight wearable sensing setup that captures eating-related bone-conducted audio through the bone-conduction unit of commodity earables and augments it with complementary reference modalities, including a laryngophone, bilateral wrist IMUs, and cameras, to support multimodal analysis and reliable annotation, as shown in Figure 1. Concretely, participants wear Honor Earbuds 3 Pro in both ears, whose built-in bone-conduction (BC) sensors capture chewing-induced vibrations within the ear canal. To establish high-fidelity reference data for ingestive events, we synchronously record laryngophone audio using a DZMIHKS6550 throat microphone, IMU signals from both

Dataset	Modalities	Ground truth	Setting	Participants	Events	Annotation Level
ACE [9]	Head + bilateral wrist motion Air-conducted earbud audio Video	Video-based annotation	Lab	6	17,080 chews 1,422 swallows 1,492 food intakes 329 drink intakes	Fine-grained event-level
Clemson [16]	Dominant-wrist motion (accelerometer + gyroscope) Ceiling-mounted video	Video-based annotation	Free-living (cafeteria)	271	24,088 bites 374 food/beverage items	Fine-grained event-level Bite timestamps with food/hand/utensil/container labels
FIC [6]	Dominant-wrist IMU (triaxial acceleration + orientation velocity) Frontal video	Video-based annotation with synchronized IMU reference	Semi-controlled environment	12	21 meals 1,332 intake cycles	Fine-grained event-level Intake-cycle intervals + wrist micromovement labels
OREBA-DIS [14]	Frontal video IMU (both wrists)	Video-based annotation	Lab (controlled meal)	100	100 recordings 4,790 intake gestures	Fine-grained event-level intake gestures (eat/drink)
OREBA-SHA [14]	Frontal video IMU (both wrists) Scales (communal dishes)	Video-based annotation	Lab (controlled communal meal)	102	102 recordings 4,279 intake gestures	Fine-grained event-level intake and serve gestures
EatSense [12]	RGB-D video Upper-body skeleton (3D poses)	Manual dense labels (video-based ground truth)	Uncontrolled dining setting	27	135 video sequences 16 eating sub-actions	Dense frame-level atomic/sub-action labels
Our dataset	Bone-conduction audio Laryngophone audio Bilateral wrist IMUs Egocentric + frontal video	Video-based annotation with synchronized multimodal reference	Lab + Free-living	18	10.07 hours total 45,582 chews 2,150 swallows	Fine-grained event-level

Table 1: Comparison of publicly available ingestive sensing datasets and our dataset. Our dataset uniquely combines commodity bone-conduction earables, laryngophone reference, and synchronized lab and free-living recordings with fine-grained chewing and swallowing annotations.

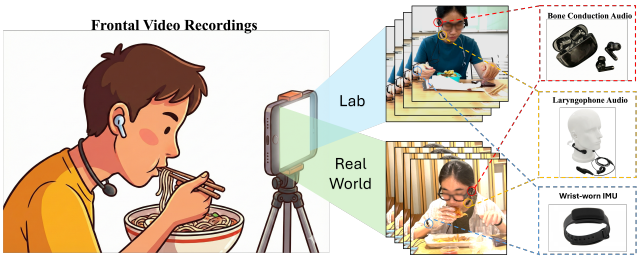

Figure 1: Multimodal data collection setup across in-lab and free-living meals. A frontal camera records the meal while participants wear commodity earbuds to capture in-ear bone-conduction audio, a laryngophone to record throat audio, a wrist-worn IMU and a first-person camera. The right panel illustrates synchronized recordings collected in both contexts and the corresponding sensing modalities except cameras.

wrists using an OPPO Watch, egocentric video using a Pupil Labs Invisible camera, and a frontal video stream captured by an iPhone 12 Pro positioned to minimize interference with natural eating, which serves as the primary ground-truth reference for annotation.

All audio streams are recorded at 16 kHz with 16-bit resolution, and videos are captured at 1080p and 30 fps. This set of modalities, including BC audio, laryngophone audio, wrist IMUs, and multi-view video, balances ecological validity and annotation fidelity. BC audio provides an unobtrusive ear-worn signal, wrist IMUs preserve the feasibility of analyzing eating-related limb movements, and laryngophone audio together with the two video views provides complementary cues for resolving subtle boundaries between swallowing and chewing during labeling, while the frontal (eye-level) view serves as the primary ground-truth reference. To enable precise cross-modal comparisons and event-level annotation, we synchronize all recordings so that ear-canal audio, throat audio, wrist IMUs, and multi-view video are accurately aligned on a shared timeline. At the start of each session, participants perform a simple synchronization gesture by double-tapping the earbuds on the table, which produces a salient time marker that is used to align all streams to a common reference point.

3.2 Procedures

This study is approved by the institutional review board, and all participants provide written informed consent before any procedures. We recruit 18 participants (9 female and 9 male), aged 20-29 (mean= 23.83, SD= 2.66). All participants are healthy adults with no history of gastrointestinal or metabolic disease, and none reports pain or discomfort during eating. Each participant completes two sessions in complementary contexts, including a quiet in-lab conference room session and an in-the-wild campus cafeteria session selected by the participant. To maximize ecological validity, sessions take place during regular meal times, participants freely choose their foods, and participants are instructed to maintain their natural eating pace and avoid intentionally controlling chewing or swallowing so that the recordings reflect everyday habits rather than staged behavior.

In the in-lab session, researchers assist with device placement before recording begins. Participants wear earbuds for dual-channel BC audio recording, a throat microphone on the neck as a reference, an egocentric camera for first-person video, and an iPhone positioned at eye level to record the meal while minimizing interference with natural behavior. After the synchronization gesture, participants proceed with eating. To complement the controlled laboratory setting, we also collect data in the cafeteria context. Each participant chooses a campus cafeteria and records a regular meal during typical dining hours with self-selected foods. The same sensing configuration is used, researchers assist with wearing the devices before the session, and video recording is conducted as unobtrusively as possible to reduce observer effects on eating behavior. This combined in-lab and free-living design maintains a consistent sensing configuration across contexts while capturing realistic variation in food types, posture, ambient sound, and social situations.

3.3 Annotation Workflow

Each recording session begins with a brief synchronization cue in which participants double-tap the earbuds on the table, producing a salient marker that aligns bone-conduction audio, laryngophone audio, wrist IMU, and both egocentric and frontal video on a shared timeline. After alignment, annotation primarily relies on the frontal video as the ground truth reference. We annotate fine-grained ingestive events using Praat and ELAN. To improve label reliability, especially for events that can be ambiguous in a single modality, we additionally consult the synchronized laryngophone audio and the egocentric video during labeling. For each session, annotators mark time intervals corresponding to chewing and swallowing events, producing event-level labels that support downstream analyses.

3.4 Dataset Summary

The dataset contains 10.07 hours of synchronized recordings, including 5.65 hours collected in the laboratory and 4.42 hours collected in free-living settings. Across these recordings, we label 24,521 (in-lab) and 21,061 (free-living) chewing events, as well as 1,114 (in-lab) and 1,036 (free-living) swallowing events. Table 1 summarizes dataset statistics and positions our dataset with respect to prior ingestive sensing datasets.

4 Preliminary Findings

Sensor	Model	Window/Hop	Accuracy	Macro-F1	F1 Chew	F1 Swallow
Laryngophone	Bagging Trees	0.5s/0.1s	0.78	0.66	0.85	0.52
Laryngophone	Random Forest	0.5s/0.1s	0.78	0.68	0.85	0.54
Bone-conducted earbuds	Bagging Trees	0.5s/0.1s	0.81	0.61	0.88	0.29
Bone-conducted earbuds	Random Forest	0.5s/0.1s	0.81	0.62	0.88	0.31

Table 2: Segment-level 3-class baselines (chew/swallow/other) on laryngophone and bone-conduction earbud signals. We report accuracy, macro-F1, and per-class F1.

To characterize the current capability of the dataset rather than to introduce a new algorithmic contribution, we report concise baselines for a three-class classification task that distinguishes chewing, swallowing, and other segments. Table 2 summarizes the best-performing tree-based models on the laryngophone and bone-conduction earbud signals.

The results reveal a clear modality trade-off. On laryngophone audio, Random Forest achieves 0.78 accuracy and 0.68 macro-F1, with strong chewing recognition (F1 = 0.85) and moderate swallowing performance (F1 = 0.54). Bagging Trees yields comparable results, reaching 0.78 accuracy and 0.66 macro-F1. In contrast, bone-conduction earbud audio supports similarly strong chewing recognition but weaker swallowing recognition. Under the best-performing segmentation setting (0.5 s window and 0.1 s hop), Random Forest reaches 0.81 accuracy and 0.62 macro-F1, with F1 scores of 0.88 for chewing and 0.31 for swallowing, while Bagging Trees obtains a similar result with a swallowing F1 of 0.29. This pattern indicates that bone-conduction earbud audio is better suited for chewing recognition, whereas laryngophone audio provides stronger cues for swallowing. This distinction is consistent with the underlying sensing mechanisms. Chewing produces prominent vibration impulses from mandibular opening and closing that propagate through cranial and mandibular bone conduction and can be captured reliably in the ear canal. Swallowing is dominated by contractions in the neck and throat, and the longer propagation path to the ear reduces the strength and separability of the corresponding signals, which makes throat-mounted sensing more effective.

We further observe that performance on the earbud channel is highly sensitive to temporal segmentation. When using a 0.5 s window and a 0.1 s hop, swallowing recognition reaches its strongest performance, with F1 scores between 0.29 and 0.31. However, when the window is reduced to 0.1 s with a 0.05 s shift, the swallowing F1 drops to 0.05 for both Bagging Trees and Random Forest. With an even shorter 0.02 s window and a 0 s shift, the swallowing F1 further decreases to 0.02 and 0.03, respectively. This pattern suggests a clear trade-off between temporal responsiveness and reliable fine-grained swallowing recognition in earable sensing.

Overall, these preliminary results show that the dataset already supports reliable chewing recognition and meaningful cross-modality benchmarking, while also highlighting two persistent challenges for future work, namely robust swallowing detection and stronger generalization across users and contexts.

5 Conclusion & Discussion

We present an aligned and annotated multimodal dataset for fine-grained ingestive sensing across laboratory and free-living meals. The dataset combines commodity earable bone-conduction audio with complementary reference modalities, including laryngophone audio, bilateral wrist IMUs, and egocentric and frontal video, and provides event-level labels for chewing and swallowing. This design supports both practical earable-centered modeling and careful inspection of failure modes through synchronized references, while covering realistic variability introduced by free-living dining.

Our preliminary benchmarks suggest that while chewing recognition can be reliably detected, swallowing remains substantially more challenging. The results also indicate that swallowing performance on earable signals is sensitive to temporal segmentation, which points to a fundamental trade-off between responsiveness and reliable fine-grained recognition in real-time settings. Together, these observations motivate future work on more robust swallowing detection, improved temporal modeling, and principled fusion of earable and reference modalities that can better handle noise, motion, and contextual variability in everyday meals.

This dataset has several limitations that should guide future extensions. Our participants are limited in number and demographic range, and the collection is constrained to two on-campus contexts. Video recordings introduce privacy considerations that may restrict release in raw form, which motivates privacy-preserving sharing strategies and evaluation protocols. In addition, bone-conduction measurements depend on fit and device characteristics, and generalization across earable hardware remains an open question. Despite these limitations, we expect this dataset to provide a useful foundation for benchmarking fine-grained ingestive sensing and for developing models and sensing designs that move beyond controlled settings toward reliable, everyday deployment.

References

[1] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. 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*. 1–12.

[2] Ai Ting Goh, Jie Ying Michelle Choy, Xin Hui Chua, Shalini Ponnalagu, Chin Meng Khoo, Clare Whitton, Rob Martinus van Dam, and Ciarán Gerard Forde. 2021. Increased oral processing and a slower eating rate increase glycaemic, insulin and satiety responses to a mixed meal tolerance test. *European Journal of nutrition* 60, 5 (2021), 2719–2733.

[3] Emil Jovanov, Edward Sazonov, and Carmen Poon. 2014. Sensors and systems for obesity care and research. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3188–3191.

[4] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. 2017. Food intake detection from inertial sensors using lstm networks. In *International Conference on Image Analysis and Processing*. Springer, 411–418.

[5] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. 2019. Detecting Meals In the Wild Using the Inertial Data of a Typical Smartwatch. In *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

[6] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. 2019. Modeling Wrist Micromovements to Measure In-Meal Eating Behavior from Inertial Sensor Data. *IEEE journal of biomedical and health informatics* (2019).

[7] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. 2020. A Data Driven End-to-end Approach for In-the-wild Monitoring of Eating Behavior Using Smartwatches. *IEEE Journal of Biomedical and Health Informatics* (2020).

[8] Matias Laporte, Daniele Gasparini, Martin Gjoreski, and Marc Langheinrich. 2022. Exploring LAUREATE-the Longitudinal multimodAl stUdent expeRience datasEt for AffecT and mEmory research. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 494–499.

[9] Christopher A Merck, Christina Maher, Mark Mirtchouk, Min Zheng, Yuxiao Huang, and Samantha Kleinberg. 2016. Multimodality sensing for eating recognition.. In *PervasiveHealth*. 130–137.

[10] AE Mesas, M Muñoz-Pareja, E López-García, and F Rodríguez-Artalejo. 2012. Selected eating behaviours and excess body weight: a systematic review. *Obesity Reviews* 13, 2 (2012), 106–135.

[11] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg. 2017. Recognizing eating from body-worn sensors: Combining free-living and laboratory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

[12] Muhammad Ahmed Raza, Longfei Chen, Li Nanbo, and Robert B Fisher. 2023. EatSense: Human centric, action recognition and localization dataset for understanding eating behaviors and quality of motion assessment. *Image and Vision Computing* 137 (2023), 104762.

[13] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 3 (2022), 1–57.

[14] Philipp V Rouast, Hamid Heydarian, Marc TP Adam, and Megan E Rollo. 2020. Oreba: A dataset for objectively recognizing eating behavior and associated intake. *IEEE Access* 8 (2020), 181955–181963.

[15] Alfonso Sánchez-Ayala, Arcelino Farias-Neto, Nara Hellen Campanha, and Renata Cunha Matheus Rodrigues Garcia. 2013. Relationship between chewing rate and masticatory performance. *CRANIO®* 31, 2 (2013), 118–122.

[16] Yiru Shen, James Salley, Eric Muth, and Adam Hoover. 2016. Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables. *IEEE journal of biomedical and health informatics* 21, 3 (2016), 599–606.

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

[18] Bryan Pak-Hei So, Tim Tin-Chun Chan, Liangchao Liu, Calvin Chi-Kong Yip, Hyo-Jung Lim, Wing-Kai Lam, Duo Wai-Chi Wong, Daphne Sze Ki Cheung, and James Chung-Wai Cheung. 2022. Swallow detection with acoustics and accelerometric-based wearable technology: a scoping review. *International journal of environmental research and public health* 20, 1 (2022), 170.

[19] MA Tuğtekin Turan and Engin Erzin. 2018. Detection of food intake events from throat microphone recordings using convolutional neural networks. In *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 1–6.

[20] Shibo Zhang, Yuqi Zhao, Dzung Tri Nguyen, Runsheng Xu, Sougata Sen, Josiah Hester, and Nabil Alshurafa. 2020. Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 2 (2020), 1–26.