Measuring children's eating behavior with a wearable device

Shengjie Bi*, Yiyang Lu†, Nicole Tobias‡, Ella Ryan*, Travis Masterson*,
Sougata Sen§, Ryan Halter*, Jacob Sorber‡, Diane Gilbert-Diamond*, David Kotz*
*Dartmouth College, Hanover, NH 03755, USA
†University of Electronic Science and Technology of China, Chengdu, China
‡Clemson University, Clemson, SC 29634, USA
§Northwestern University, Evanston, IL 60208, USA
Email: shengjie.bi.gr@dartmouth.edu

2020 IEEE International Conference on Healthcare Informatics (ICHI) | 978-1-7281-5382-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICHI48887.2020.9374304

***Abstract*—Poor eating habits in children and teenagers can lead to obesity, eating disorders, or life-threatening health problems. Although researchers have studied children's eating behavior for decades, the research community has had limited technology to support the observation and measurement of fine-grained details of a child's eating behavior. In this paper, we present the feasibility of adapting the *Auracle*, an existing research-grade earpiece designed to automatically and unobtrusively recognize eating behavior in adults, for measuring children's eating behavior. We identified and addressed several challenges pertaining to monitoring eating behavior in children, paying particular attention to device fit and comfort. We also improved the accuracy and robustness of the eating-activity detection algorithms. We used this improved prototype in a lab study with a sample of 10 children for 60 total sessions and collected 22.3 hours of data in both meal and snack scenarios. Overall, we achieved an accuracy exceeding 85.0% and an F1 score exceeding 84.2% for eating detection with a 3-second resolution, and a 95.5% accuracy and a 95.7% F1 score for eating detection with a 1-minute resolution.**

I. Introduction

Obesity has become a serious threat to public health in America. In most cases, obesity is caused in part by over-consumption of food, so individualized feedback about eating habits may help reduce obesity rates. This information is most pertinent early in the lifespan, prior to excess weight gain and the development of obesity. Childhood obesity rates continue to be high (18.5 percent in 2016) in the United States [1] and are associated with a myriad of co-morbidities that negatively impact overall quality of life [2]. Furthermore, weight-related issues in childhood are likely to carry into adulthood [3]. It is therefore essential to improve our scientific understanding of childhood eating behaviors to inform obesity interventions. Indeed, individualized, just-in-time adaptive interventions (JITAIs) focused on eating habits may be effective in reducing over-consumption in children [4], but are not feasible until there is technology that can automatically detect and measure eating behavior.

With continued research into Automatic Dietary Monitoring (ADM) systems, monitoring fine-grained details of the eating activity is gradually becoming possible. These ADM systems have the capability of triggering real-time interventions during problematic eating activities. Existing ADM techniques still face several inherent challenges: individuals can eat at varied speeds, perform confounding actions during eating activities, and vary the type and mode for consuming different types of food. To monitor eating behavior in children, we face all those challenges and more: children usually have more non-eating related head and body movement during eating, children have more complex eating behaviour (e.g., children may hold and play with food in their mouths for a while before chewing and swallowing), children's head and body sizes vary more than adults, and children are more sensitive to the discomfort of wearable devices [5], [6]. Although several researchers have evaluated ADM systems on adults, no automatic dietary monitoring technique exists for children. Researchers and behavioral scientists depend on traditional techniques such as video coding and manual food journals to monitor dietary activities among children [7]. To better support the needs of clinicians and behavioral scientists in monitoring eating habits among children, we modified an existing ADM system that had previously been tested only among adults. In this paper, we report the insights gained and results obtained from experiments with a new child-oriented ADM system derived from the *Auracle* wearable ADM system [8], [9].

To determine the performance of this ADM system, we conducted a set of controlled experiments. During this study, the participants (children) visited our laboratory on multiple occasions and consumed a variety of meals while wearing the modified Auracle system. Our initial findings indicate that it is indeed possible to identify and monitor fine-grained eating activities of children, once we addressed specific challenges. With further refinement, we believe that such an ADM system may also be used to monitor a child's eating activity in naturalistic settings.

Accurate high-resolution eating detection could help trigger other kinds of sensing or inquiries [8]. Specifically, we believe it is important to develop ADM techniques that can detect eating (whether the user is eating or not), within a few seconds of eating onset, to enable (1) detailed analysis of eating patterns like mouthful rate, chewing rate, and consumption rate, and (2) to enable just-in-time interventions in free-living conditions. For instance, researchers have recently shown that poor mastication is associated with obesity [10]. Additionally,

978-1-7281-5382-7/20/$31.00 ©2020 IEEE

if we want to estimate the caloric intake of a meal, we may need to classify different types of food consumed during the meal, and thus require eating detection to identify the precise moment of mastication for each food item. We set out to enable such capabilities for monitoring children, and believe it to be the first effort to do so.

In this paper, we make two important contributions:

- We demonstrate that it is feasible to monitor the eating activity of children automatically. This result provides the foundation for behavioral researchers, clinicians, and dietitians to understand fine-grained details about a child's eating habits. In both meal and snack scenarios involving 10 children over a total of 60 lab sessions, we achieved an accuracy exceeding 85.0% and an F1 score exceeding 84.2% for eating detection with a 3-second resolution. The same methods obtained a 95.5% accuracy and a 95.7% F1 score for eating detection with a 1-minute resolution.
- We identify unique challenges pertaining to the use of existing ADM systems (designed for an adult population) on children. We detail the steps necessary to adapt an adult device to allow data collection from children. With these adaptations, we developed the first ADM system focused on the study of eating behavior in children.

II. Background

First, some background information about ADM systems.

a) Defining eating: In our prior work, we defined *eating* as "an activity involving the chewing of food that is eventually swallowed" [9]. This definition will exclude drinking actions because drinking usually does not not involve chewing. The definition will also exclude chewing gum as that usually does not involve swallowing. Additionally, we defined an *eating episode* as: "a period of time beginning and ending with eating activity, with no internal long gaps, but separated from each adjacent eating episode by a long gap, where a *gap* is a period in which no eating activity occurs, and where *long* means a duration greater than a parameter δ" [9]. We chose $\delta = 15$ minutes, drawing on precedent in Leech et al. [11].

b) The Auracle ADM system: In this work, we used our Auracle ADM system. Auracle is a head-mounted device with a form factor similar to a behind-the-head pair of earphones (Figure 1); we believe a professionally-engineered version of this design would be small, safe, and comfortable for a child to wear. This design places a skin-contact microphone behind the wearer's ear, to capture the sound of a person chewing; this approach should be safer than placing a microphone or other sensor in the ear canal, and less disruptive to normal hearing. Since the Auracle is out of view of the child, we speculate that it might be less distracting than anything worn on the top or front of the head. Nonetheless, for this paper we developed a new approach (details in Section IV) that we believe is an even more natural choice.

In our previous work, we found this device could automatically recognize *when* and for *how long* a person is eating: we collected field data with 14 adult participants for 32 hours,

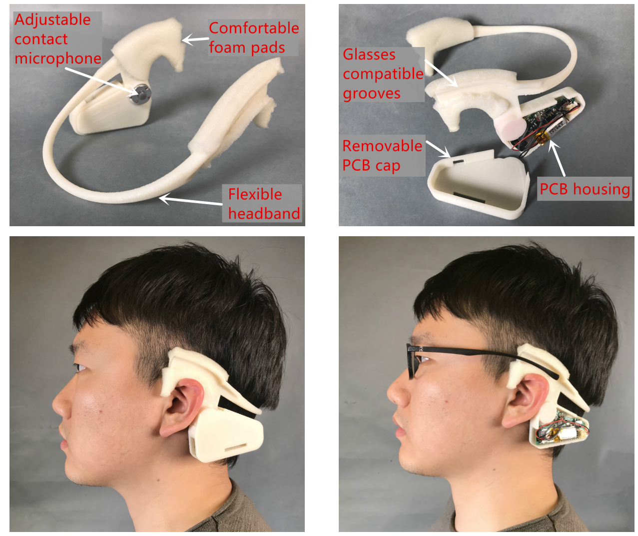

Fig. 1. Original mechanical housing of Auracle; figure from [9].

and achieved F1 score exceeding 77.5% for eating detection in free-living conditions [9].

Furthermore, we estimated Auracle's battery should last up to 28 hours with a 110 mAh battery while continuously monitoring eating and communicating its observations to a smartphone over Bluetooth [9]. Although the battery life or communication capabilities are not critical for the in-lab studies we conducted for this paper, these system capabilities and performance criteria are important for our envisioned usage of the ADM system in naturalistic settings.

III. Related work

Researchers have long been interested in measuring various parameters of eating behavior, such as time, duration, chew rate, bite rate, and meal-specific data such as food quantity, food-group classification, and calorie estimation. For all of these parameters, accurate recognition of *when* people eat is the foundation of effective automatic dietary monitoring (ADM) systems.

Researchers have developed ADM systems that use various cues for eating detection, including audio collected from the ear canal [12]–[18], behind the ear [8], [9], [19] or on the throat [14], [20]–[23], proximity of a necklace from the chin [24], first-person images from chest-mounted cameras [25]–[27], or wrist-based gesture recognition [28], [29]. Each approach has been tested on adults and has its own limitations and advantages. To select a suitable device for children, we aimed to avoid devices that could be dangerous in free-living conditions (tiny microphone in the ear canal), privacy invasive (images capturing the child, or other children), socially unacceptable (device on throat), or easily distracting during regular use (such as a wristband on the dominant hand or on both hands). Furthermore, any ADM system aimed at free-living conditions must be accurate, compact, light, comfortable, cheap, robust, usable, and energy efficient.

Our work bridges the gap between two types of research: studies that focus on understanding children's eating behavior, and studies that use wearable sensors for automatic eating detection.

A. Children's eating studies in health science

Eating behavior of children has long been a important topic in health-science research. Researchers have studied various issues related to *when* and *how long* children eat. For instance, Klesges et al. found that time spent eating in a meal correlates to weight, but not to the total meal time (i.e., time spent at the table) [30]. In studies pertaining to monitoring children's eating habits, researchers depend on traditional techniques such as video coding and manual bookkeeping for recognizing when and how long children eat. By developing a wearable sensor that can accurately detect eating, we believe most of these studies could be completed with finer granularity, higher accuracy, and substantially less labor.

Another common way to assess children's eating behavior is through the evaluation of the eating micro structure, including aspects such as bite size, eating rate (bites/minute) and meal length. For instance, Llewwllyn et al. have shown that children with higher eating rate tend to have higher body weight [31]. Accurate recognition of *when* children eat is the foundation of ADM systems that help collect micro-structure eating information for health-science researchers. A wearable system usable in free-living settings could capture metrics about eating outside mealtimes, which can influence a child's health but would be difficult to measure with traditional methods.

B. Eating detection with wearable sensors

A comprehensive survey of this topic is not feasible here; we focus only on some of the most recent work, and direct the interested reader to survey papers that cover aspects of eating detection and ADM systems [32]–[36]. Sen et al. built and tested an approach based on wrist motion and achieved false-positive and false-negative rates of 6.5% and 3.3% respectively [28], [37]. Mirtchouk et al. experimented with different combinations of motion (head, wrist) and audio (air microphone) data collected in laboratory and free-living conditions [38]. They found a combination of sensing modalities (audio, motion) was needed, but sensor placement (head vs. wrist) was not critical. Zhang et al. proposed and evaluated smart eyeglasses using electromyographic (EMG) sensors and achieved precision and recall more than 77% for chewing detection in free-living scenarios [39]. Bedri et al. evaluated optical, inertial and acoustic sensors, and selected a behind-the-ear inertial sensor; they achieved an F1 score of 80.1% for detecting eating episodes in the field study [40]. Using a proximity sensor, Chun et al. developed a necklace that captures head and jawbone movement, and achieved 78.2% precision and 72.5% recall for detecting eating episodes in a free-living study [24]. Zhang et al. developed a low-power necklace and achieved a F1 score of 77.1% for detecting eating episodes in an all-day free-living setting [41]. Additionally, to help with behavior change, some researchers paid particular attention to just-in-time eating intervention [42], [43].

All of the above systems were evaluated only with an adult population. In contrast, we developed the first ADM system that can monitor children's eating habits, and evaluated our system in both meal and snack scenarios.

IV. System design

The Auracle system includes a contact microphone, a battery, a custom-designed printed circuit board (PCB) for data acquisition and a wearable mechanical housing (Figure 1). Since the device was primarily designed for data collection with adults, we had to modify the housing of the device to ensure that it performed reliably in detecting children's eating activities while ensuring that it did not distract or discomfort the child.

For this study, we updated both the hardware and software of Auracle. The updated PCB had several new or improved features relative to the version that we described in the original paper [9]. These updates included replacement of the original Texas Instruments (TI) CC2640R2F microcontroller with the MSP430FR5994 microcontroller, addition of a new BLE chipset (Nordic nRF51822), and addition of an accelerometer (ADXL362). We did not use the accelerometer or BLE communication in our current study. We used the Auracle to collect 10-bit samples of the microphone signal at 500 Hz and write the data to the SD card. In this study, the most beneficial aspect of the updated Auracle hardware was that the total size of the board was reduced by over 50% (Figure 2): now smaller than 37×22mm; it was easier to use this board to design a device suitable for the smaller heads of children.

Based on our preliminary tests, we observed that children's heads vary tremendously in size and shape, making it necessary to design a form factor that could easily adapt to a range of children. Although the Auracle's contact microphone's position and the pressure that it applied to the skin were adjustable in the original design, preliminary testing showed that it did not provide adequate contact for several children, rendering the collected data inadequate for analysis. This observation prompted us to house the Auracle in an elastic headband (Figure 3) rather than in the original 3d-printed plastic frame (Figure 1). The elastic headband ensured that the device was comfortable and robust to movement, and the microphone maintained proper contact with the child's skin. It also adapted to a wide range in head sizes without requiring any mechanical design modification. Furthermore, it was less distracting for the child during the in-lab studies.

V. Data collection

We trained several research assistants to use our modified Auracle in this study, following a protocol approved by our Institutional Review Board (IRB).

A. Laboratory Data Collection

We collected data from 10 children (aged 4-17; 4 female, 6 male). Each participant visited the lab on 3 occasions and we

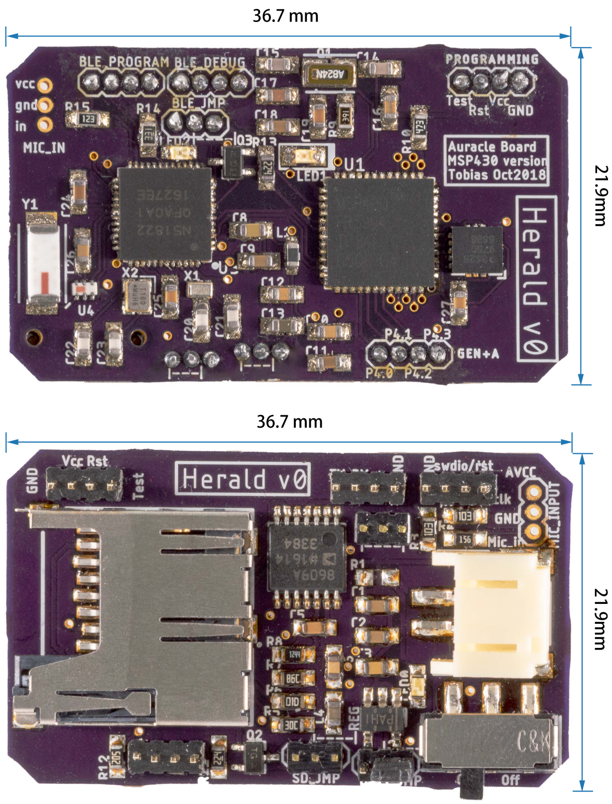

Fig. 2. The top and bottom view of the updated and improved Auracle's printed circuit board (PCB).

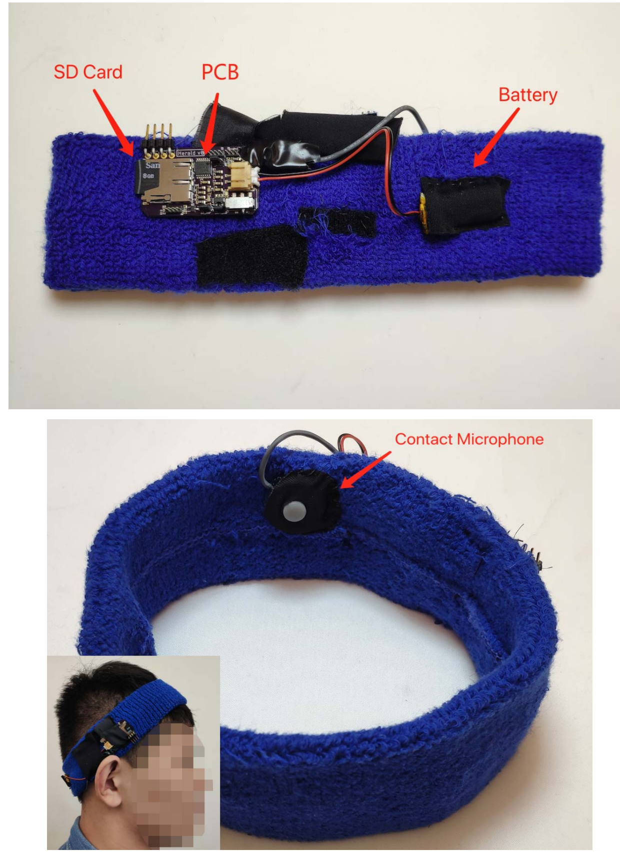

Fig. 3. Auracle prototype after our revision, using elastic headband.

collected data from two sessions per visit: one meal and one snack. Overall, our dataset consists of 30 meal and 30 snack sessions. After a preliminary review of the data, we determined that we could not use data from 16 sessions, collected from 4 different participants, for the following reasons. In four of these sessions, the contact microphone signal was weak due to poor contact or improper placement of the microphone, and the signal barely changed during these sessions. In the other twelve sessions, the data was not usable because research assistants forgot or incorrectly performed some of the procedures in our protocol (e.g., turn on the camera, start with three-tap event) or because participants inadvertently interfered with the Auracle (e.g., touching the headband frequently). For our final analysis, we excluded the data collected during these 16 sessions. We used the remaining 44 sessions of recorded data (16.86 hours in total) from 8 participants for further analysis.

B. Data Collection Protocol

At the start of each session, a research assistant placed the Auracle device around the participant's head and adjusted the contact microphone so that it was located on the mastoid tip, behind the ear. The participants were instructed not to adjust or remove the Auracle device during the study. We placed a Go Pro camera in front of the participants to film their eating behavior. We later annotated the videos to provide 'ground truth' about the participants' eating behavior.

We first conducted the 'meal' session, in which we served pre-determined portions of three food items to participants (macaroni and cheese, carrots, and apple slices). Participants sat in front of a dining table during the meal, and we encouraged them to perform the eating activity as they normally would in a naturalistic setting. Participants were provided additional servings of any food type if they completed the initial serving and indicated that they wanted more of that food type. After a short break, the 'snack' session began: we provided another three food types (gummy bears, grapes, and goldfish crackers) to the participants. Participants sat on a sofa, in front of a TV, watching a show (with commercials about food) for 30 minutes. An example of one session of data collection is shown in Figure 4. The red portions in the figure were human-annotated to indicate eating periods. Figure 5 shows two screen shots of the video recorded by the camera from two participants during the meal and snack sessions, respectively. In general, we found that participants were more relaxed and natural in the snack session than the meal session. Overall, none of the participants complained about any discomfort caused by the device and did not remove it during their sessions.

At the beginning of each session, we asked the participant to simultaneously tap on their cheek and on the headband three times using their hand, while ensuring that the camera could record this action. At the end of the session, we asked the participant to again perform this 'triple-tap' action. We

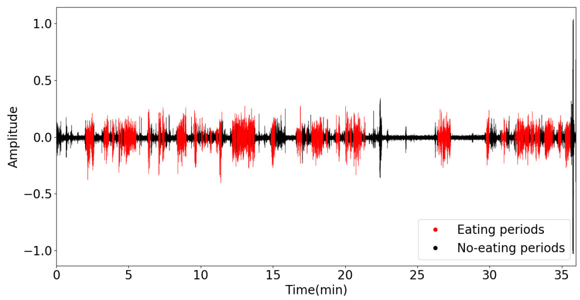

Fig. 4. Temporal signature of one session of data collection (red portions indicate periods of eating).

Fig. 5. Screen shots of the video recorded during meal and snack sessions.

later identified these triple-tap events in the video (from the camera) and the audio (from the microphone) and used them to synchronize the video and audio data streams.

During the data collection, at least one research assistant was always present in the room with the child for safety reasons. The research assistant visually checked the position of the headband periodically to ensure the device stayed at the proper location during the study. However, the research assistant pretended to focus on paperwork, and avoided talking to or distracting the participants. We also asked one parent of the child to wait near the laboratory, to address any unexpected situations. We compensated each participant with a \$30, \$35, or \$40 gift card for the first, second, and third visits, respectively.

C. Video Annotation

We used a commercial service to annotate the videos.[1] The annotation process consists of three steps: execution, audit, and quality inspection. In the *execution* step, an annotator watched the video and annotated each period of eating, at a 1-second resolution. Thus, for every second in the video, the annotator indicated whether the child was eating or not. Next, in the *audit* step, an auditor watched the video and checked whether the annotations were consistent with the content in the video. The auditor noted any identified inconsistency for the next step: quality inspection. Finally, in the third step, a *quality inspector* reviewed the questionable labels and made the final decision about each identified inconsistency. The *quality inspector* also conducted a second-round inspection of 20% of the samples that were considered consistent during the previous two inspection rounds. This three-phase, three-person process ensured that the quality of the video annotation was acceptable.

VI. Data analysis

We next describe our evaluation metrics, and the stages of our data-processing pipeline: preprocessing, feature extraction, classification, and aggregation.

A. Evaluation Metrics

We set out to evaluate our method for fine-grained eating detection *(window-based classification)* and for coarse-grained eating detection *(episode-based classification)*, as detailed in the subsections below. Since we aim for generalized models, we use a 'Leave-One-Session-Out' (LOSO) approach to evaluate model efficacy. In a LOSO approach, data from one session of a participant is tested on a model that has been trained using a combination of data from all other sessions of the same participant and every session of all other participants. Formally, if the dataset has data from I participants, each of whom has provided data for J sessions, then set S_{ij} represents participant i's data from session j, for $i \in \{1, 2, \cdots I\}$ and $j \in \{1, 2, \cdots J\}$. Overall, set $S = \cup_{\forall i,j} S_{ij}$ represents all sessions in the dataset. Then the model is trained using sessions in the set $S - S_{ij}$ and tested on session S_{ij}. This process is repeated so that every session of every participant is tested on a model generated from all sessions in the dataset, except the session being tested.

In preliminary tests, we observed that the data we collected from a participant in different sessions could often vary in signal amplitude. One reason for this difference is because the same participant might wear the Auracle device differently (e.g., the angle of wearing the headband) during different sessions, which caused the contact microphone to be located at different locations or in contact with the skin with different pressure. Moreover, actions during the session (such as touching the device during the session or scratching the head) may also have affected the microphone contact. Thus, we first applied the normalization approach mentioned in Section VI-B, and then chose a LOSO cross-validation approach to test the performance

[1] BasicFinder: https://www.basicfinder.com/en/

of the classifier in detecting the eating activity for data in a session that it has never seen before.

1) LOSO Window-based Evaluation: In window-based evaluation, we explored two window sizes: 3 seconds and 1 minute. Three-second windows are important for applications that rely on the output of ADM systems to drive fine-grained interventions (e.g., an in-the-moment intervention based on the mastication habit). One-minute windows enable us to compare our results with results presented in our prior study [9]. For each window size, we compare our classifier's output against the ground-truth label for the corresponding time window, computing four evaluation metrics (accuracy, precision, recall, and F1 score) for each session, then averaging those metrics across sessions to compute the four summary metrics for that window size. We used the same metrics as the evaluation in our previous work [9].

2) LOSO Episode-based Evaluation: In episode-based evaluation, using an approach similar to previous work by Papapanagioto et al. [44], we matched each detected eating episode with either 0 or 1 ground-truth eating episode. We used the Jaccard Similarity coefficient to determine whether this match led to a *Correct Detection*, *False Detection*, or *Missed Detection*. Our definition of the Jaccard Similarity coefficient is the same as in our previous work [9], as follows.

> Let the detected episode be represented as $E_d = [t_s, t_e]$, where t_s is the start of the detected eating episode and t_e is the end of the detected eating episode. Similarly, the actual eating episode (obtained from ground truth) is represented by $E_a = [t'_s, t'_e]$, where t'_s is the start of the actual eating episode and t'_e is the end of the actual eating episode.
> We then compute the Jaccard similarity coefficient:
>
> $$J = \frac{E_a \ \cap \ E_d}{E_a \ \cup \ E_d}$$
>
> Each detected eating episode is an independent test case that results in one of three outcomes:

$$Outcome = \begin{cases} J \geq 0.55, & \text{Correct Detection} \\ 0 < J < 0.55, & \text{False Detection} \\ J = 0, & \text{Missed Detection} \end{cases}$$

We then count the number of Correct Detections, False Detections, and Missed Detections.

B. Data Processing Pipeline

Figure 6 presents our overall data-processing pipeline, which comprises preprocessing, feature extraction, classification, and aggregation steps.

1) Preprocessing: The preprocessing stage includes three steps: RMSE, normalization, and segmentation. As noted above, the audio-signal amplitude can be affected by the location of the contact microphone and the pressure applied to it. We observed that the signal amplitude varied from session to session due to differences in position, pressure, and head size/shape. To ensure uniformity, we used the root mean square energy (RMSE)

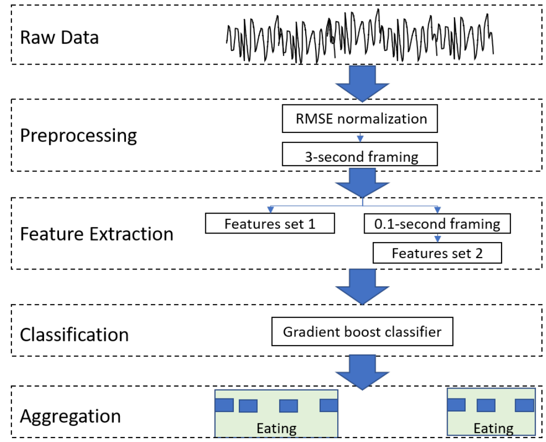

Fig. 6. Data processing pipeline.

value to normalize the signals within each session. However, we found motion artifacts in some portions of the signal; these artifacts were usually caused by movement of the contact microphone across the skin, and can have an outsized effect on the RMSE value. After some preliminary tests, we decided to exclude samples that were not within the 95% confidence interval when we calculated the RMSE value of each session's signal. Then, in the normalization step, we divided all the data values by the RMSE value of the same signal. Then, we segmented the acoustic signals into non-overlapping 3-second windows of samples, and passed these windows to the feature-extraction stage. Note: our current normalization method computes RMSE across the entire session and is thus only suitable for offline processing; one can envision similar normalization approaches suitable for online processing, e.g., normalizing each sample by dividing by the RMSE computed over a period of recent data.

2) Feature Extraction: For each time window, we extracted the 30 features shown in Table I, including 20 frequency-domain features and 10 time-domain features selected from about 1,400 possible features (see Section VI-C2 for details). We extracted some of these features directly from the windows received from the preprocessing stage, using the tsfresh[2] package. We extracted the other features using methods similar to those used by Bogdanov et al. [45], using the librosa[3] package. In this latter case, we segmented each 3-second window into 0.1-second 'frames' (with 75% overlap between adjacent frames) and extracted features from each frame. For each feature and each window, we obtained an array of values corresponding to the 0.1-second frames. We then computed eight statistics (mean, median, variance, maximum, minimum, kurtosis, skewness, entropy) for each array. Using these eight statistics of each feature array, we extracted features for each time window.

3) Classification and aggregation: The classification stage has two steps: classification and aggregation. First, the Gradient

[2]v0.11.2: http://tsfresh.readthedocs.io/en/latest/
[3]v0.7.2: https://librosa.github.io/librosa

TABLE I
30 FEATURES USED BY OUR CLASSIFIER (* INDICATES THE FREQUENCY-DOMAIN FEATURES)

Feature category	Description	Number of features	Feature set
MFCCs*	Mel-frequency cepstral coefficients	14	2
MFCCs delta*	First derivatives of MFCCs	3	2
MFCCs delta 2*	Second derivatives of MFCCs	2	2
Spectral contrast*	Difference between the spectral peak and valley in each frequency subband	1	2
Change quantiles	Mean of the absolute change of the series inside a corridor	6	1
Agg autocorrelation	Value of an aggregation function over the autocorrelation for different lags	1	1
Agg linear trend	Attributes of a linear regression for values that were aggregated over chunks	1	1
Ratio beyond r sigma	Ratio of values that are more than r*std(values) away from the mean	1	1
Quantile	Value of the data point greater than q% of the ordered values	1	1

Boosting (GB) classifier used each window's features to classify that window into one of two classes: eating or not eating. (See Section VI-C1 for details about our selection of the GB classifier). Specifically, we used the GB implementation from XGBoost.[4] Using the same aggregation methods as in this paper [9], the aggregation step combines groups of twenty 3-second windows' classification outputs into 1-minute outputs, and then further combines these 1-minute outputs into an episode-level output. We also apply the same two-step aggregation process to the ground-truth labels (which has a base resolution of 1 second).

C. Classifier and feature selection

At this point we digress to justify our choice of the GB classifier and our selection of the 30 specific features listed in Table I. To make these decisions, we conducted several benchmark studies to determine the best-performing classifier (in terms of F1 score) and the most discriminative features.

1) Choice of classifier: We initially assumed we would use the Logistic Regression (LR) classifier because we identified LR as having provided the highest F1 score in determining eating segments in our previous work [9]. We decided to revisit this selection, however, because we wanted to explore a broader range of features, and because our previous study was conducted with adult participants, consuming different food types, and in free-living conditions. It seemed plausible that a different classifier, and different feature set, would be better suited for eating detection in children, or in lab settings.

[4] v0.9.0: https://xgboost.readthedocs.io/en/latest/python

TABLE II
CLASSIFIER PERFORMANCE WHEN USING TOP-30 FEATURES FROM ONLY FEATURE SET 1

Classifier	Accuracy	Precision	Recall	F1 score
Gradient boosting	0.819	0.810	0.839	0.815
Random forest	0.816	0.815	0.820	0.809
K-nearest neighbors (K=5)	0.802	0.793	0.814	0.796
Logistic regression	0.793	0.818	0.776	0.786
Support Vector Machine	0.813	0.813	0.818	0.807
Gaussian Naive Bayes	0.802	0.760	0.884	0.809

TABLE III
CLASSIFIER PERFORMANCE WHEN USING TOP-30 FEATURES FROM BOTH FEATURE SETS 1 AND 2

Classifier	Accuracy	Precision	Recall	F1 score
Gradient boosting	0.850	0.834	0.869	0.842
Random forest	0.845	0.841	0.842	0.833
K-nearest neighbors (K=5)	0.823	0.805	0.847	0.818
Logistic regression	0.829	0.839	0.828	0.822
Support Vector Machine	0.840	0.832	0.851	0.832
Gaussian Naive Bayes	0.825	0.820	0.829	0.815

We began with a large set of 750 features extracted with tsfresh; let that be called *feature set 1*. After inspecting these features, we found many of them are constant numbers and not useful for classification. We also anticipated it was not necessary to use all the features based on the analysis in our previous work [8], [9]. Due to these two reasons, we decided to select the best classifier when using a smaller number of features. (More discussion about feature selection can be found in Section VI-C2.) We then ran our entire dataset through our data pipeline, using only the top-30 features selected from feature set 1, using each of six common classifiers, resulting in the metrics shown in Table II. (We found adding more features yielded little-to-no improvement to F1 scores, across these six classifiers.) In Figure 8 and Figure 9, we use the GB classifier as an example to show the performance of our model when top k features were used ($1 \leq k \leq 60$).

For further confirmation, we added a set of 650 features extracted with librosa (as described above); let that be called *feature set 2*. We again ran our entire dataset through our data pipeline, using the top-30 features selected from both feature set 1 and 2, using the same six classifiers, resulting in the metrics shown in Table III. All six classifiers achieved a better F1 score relative to Table II (average improvement 2.3%, with a p-value of 0.0007), indicating that features in feature set 2 were indeed useful in the classification process.

Finally, for deeper insight into the differences among the classifiers, we plotted the Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) in Figure 7 for all classifiers; it displays the relationship between the true-positive rate and the false-positive rate of our models.

Based on results in Table II, Table III and Figure 7, GB, Random Forest, and Support Vector Machine outperformed the other three classifiers. Although the best three classifiers had similar performance, GB was slightly better so we selected

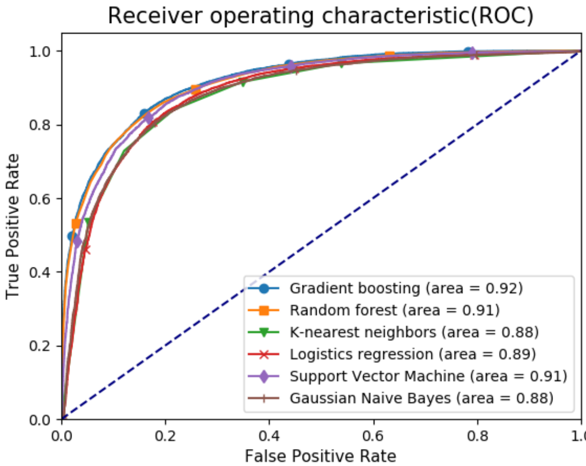

Fig. 7. ROC curve for various classifiers.

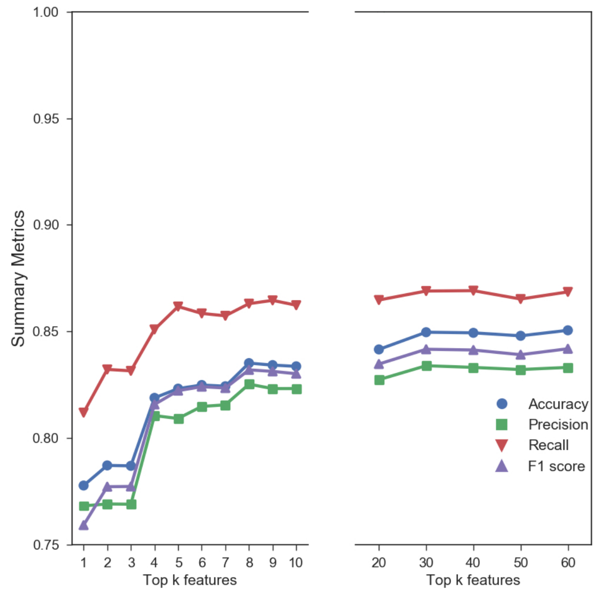

Fig. 8. Performance of the GB classifier with 3-second resolution.

GB for our analysis.

2) Feature selection: To determine how many features to use, and which features to use, we began by computing about 1,400 features (feature sets 1 and 2).

For feature set 1, we extracted 62 categories of common features directly from windows produced by the preprocessing stage; 4 frequency-domain categories and 58 time-domain categories. In each feature 'category', we extracted features with all possible parameters. Some feature 'categories' can result in hundreds of features by varying the category's parameters. In our case, feature set 1 consisted of about 750 features.

For feature set 2, we extracted 14 categories of frequency-domain features. Again, some feature categories can result in hundreds of features by varying the category's parameters. In our case, feature set 2 consisted of about 650 features.

Clearly, it would be too complex to compute all 1,400 features from these two feature sets, on small wearable platforms, so we used the Recursive Feature Elimination (RFE) algorithm to identify the subset of features that were most 'discriminative'. That is, we ran our entire pipeline (with each classifier) over the complete dataset, letting RFE empirically identify the subset of k features that were most useful in distinguishing eating from non-eating moments (for each classifier). As an example, Figure 8 shows the performance of the GB classifier in eating detection, when the top k features were used ($1 \leq k \leq 60$), with a 3-second resolution. From the figure we can see that the performance improved until $k = 30$ and then it saturated. When experimenting with other classifiers, we found the trend of curves are similar.

To further understand the effect of k, using the aggregation method mentioned in Section VI-B3, we computed the performance of the system at a 1-minute resolution. Figure 9 shows the performance of the GB classifier in eating detection, when the top k features were used ($1 \leq k \leq 60$), with a 1-minute resolution. Interestingly, saturation at the 1-minute resolution occurs even earlier, indicating that the system can perform adequately with a small feature set. As previously indicated in Section I, one of our long-term goals is to provide interventions based on fine-grained eating-related actions, so we decided to use $k = 30$ and list the top-30 features in Table I. As it happens, 20 features out of 30 are frequency-domain features. Most of the frequency-domain features are MFCC features, and the first and second derivatives of MFCCs that were obtained from feature set 2, thus showing the usefulness of the features from the librosa package.

VII. Performance Evaluation

Overall, we evaluate how well our method worked

- for fine-grained eating detection (3-second windows),
- for medium-grained eating detection (1-minute windows), and
- for detecting eating episodes.

As noted above, we evaluated our method using LOSO cross validation; Table IV summarizes the resulting performance metrics for fine-grained eating detection (3-second windows) and for medium-grained eating detection (1-minute windows). In these experiments we used the GB classifier and the top-30 features (Table I), achieving an F1 score of 0.842 for fine-grained eating detection and 0.957 for medium-grained eating detection.

To better compare with our previous work [9], we also performed an episode-based evaluation. According to our definition of *eating episode*, there were 45 actual eating episodes in our laboratory data. The episodes ranged in duration from 1 minute to 38 minutes, with mean value 20.01 minutes and standard deviation 12.09 minutes. When using the

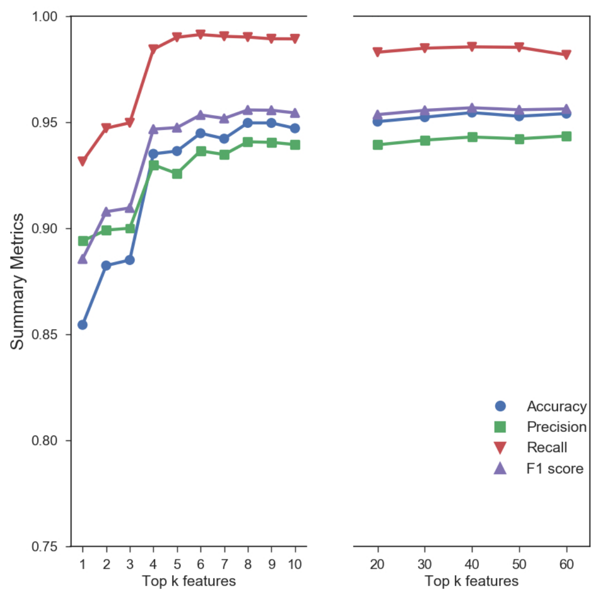

Fig. 9. Performance of the GB classifier with 1-minute resolution.

TABLE IV
METRICS FOR GB CLASSIFIER WITH TOP-30 FEATURES (MEAN VALUE ± STANDARD DEVIATION)

Training data	Session(s)	Accuracy	Precision	Recall	F1 score
3-second time window	meal	0.880 ±0.079	0.897 ± 0.094	0.923 ±0.050	0.907 ±0.063
	snack	0.820 ±0.085	0.771 ± 0.200	0.815 ±0.139	0.776 ±0.165
	both	0.850 ±0.088	0.834 ± 0.167	0.869 ±0.120	0.842 ±0.140
1-minute time window	meal	0.991 ±0.032	0.990 ± 0.037	1.000 ±0.000	0.995 ±0.020
	snack	0.918 ±0.124	0.896 ± 0.208	0.971 ±0.064	0.919 ±0.167
	both	0.955 ±0.097	0.943 ± 0.155	0.986 ±0.047	0.957 ±0.124

Jaccard similarity coefficient as our evaluation metric, and the aggregation methods mentioned in Section VI-B3, we correctly detected 43 eating episodes, missed just 2 eating episodes, and falsely detected 0 eating episodes. We next examined the difference in duration between detected eating episodes and actual eating episodes; the mean difference was only 0.76 ± 3.56 minutes. To further challenge our model, we increased the Jaccard similarity coefficient starting from 0.55 and found we were still able to achieve the same performance when the we increased the coefficient up to 0.76.

Although our method seems highly effective at detecting eating episodes, we must note that our in-lab sessions were relatively short (the longest sessions is 40.35 minutes) and participants were eating more than half of the time in typical sessions (the time-length ratio of data labeled as *non-eating* and *eating* is 0.91:1). The situation would not be challenging for any episode detector, so we cannot draw any firm conclusions about our method's ability in this regard.

VIII. Discussion and Future Work

Here we discuss our findings from the current study and ideas we may explore in the future.

a) Field studies: Our work is the first to evaluate an ADM system on children in a laboratory setting. In this paper, we show that it is indeed possible to identify children's fine-grained eating habits. It will be necessary to conduct multi-day field studies to evaluate our system's performance in free-living scenarios. Such a field study could also determine the feasibility of triggering interventions when a system identifies problematic eating habits. For a completely free-living study, our device would need the capability to determine whether the device is properly worn. Specifically, the system should notify the wearer if it detects that there is improper contact between the microphone and the skin, so the wearer can adjust the device as necessary. Such a capability will also allow the user to properly don the device after removing the device for a break. We could explore using the acoustic signal from the user's heart beat, or sounds of speech, as an indicator for proper contact after a user fits the device.

b) Placement for the contact microphone: Even in a lab setting, our research assistants sometimes found it challenging to identify the mastoid tip and the proper placement of the contact microphone. As a result, there were a few cases where the signal captured by the contact microphone was too weak for further data analysis. A revised device should perform a quantitative evaluation of the audio signal when the microphone is placed at different locations behind the ear, guiding the health-science researcher toward the best placement.

We also found that it is harder for the microphone to maintain its position for participants with long hair, especially when participants did not tie their hair back. (Four of our ten participants had long hair.) Future designs should consider means to ensure microphone contact for wearers with long hair, eyewear, or headwear.

c) Meal vs. snack: When using the top 30 features and GB classifier for eating detection with a 3-second resolution, we achieved a F1 score of 0.907 ± 0.063 in meal sessions and 0.776 ± 0.165 in snack sessions, respectively. To better understand why snack sessions were so much less accurate, we viewed the ground-truth videos to study behavior of participants in both settings. We observed that participants generally had more vigorous body movement in snack sessions (e.g., switching between sitting and lying down on sofa, putting legs onto and off of the coffee table), which in turn increased the classification error in snack sessions. The body movement was more limited when the participant sat on a chair and ate a meal that was placed on the table. In the future, it may be useful to first classify the scenarios of eating, and then use a classification model specific to that scenario.

IX. Conclusion

In this paper, we adapted *Auracle*, an existing ADM system designed for adults, and applied it in a study with children.

Indeed, we believe this paper represents the first work to develop and evaluate a wearable Automated Dietary Monitoring (ADM) system for children. Using our adapted Auracle device, we collected data with 10 participants for 60 sessions (22.3 hours) in meal and snack scenarios. We designed a data-processing pipeline and evaluated its performance using LOSO cross validation. Overall, we achieved an accuracy 85.0% and an F1 score 84.2% for eating detection with a 3-second resolution, and a 95.5% accuracy and a 95.7% F1 score for eating detection with a 1-minute resolution. Please follow us at auracle-project.org.

X. Acknowledgement

This research results from a research program at the Center for Technology and Behavioral Health (CTBH) at Dartmouth College, supported by the National Science Foundation under award numbers CNS-1565269, CNS-1835983, CNS-1565268, and CNS-1835974, and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development under award number R01HD092604. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

References

[1] C. M. Hales, M. D. Carroll, C. D. Fryar, and C. L. Ogden, "Prevalence of Obesity Among Adults and Youth: United States, 2015–2016. NCHS data brief, no 288. Hyattsville, MD: National Center for Health Statistics." *NCHS data brief, no 288. Hyattsville, MD: National Center for Health Statistics.*, 2017, Online at https://www.cdc.gov/nchs/products/databriefs/db288.htm.

[2] J. B. Schwimmer, T. M. Burwinkle, and J. W. Varni, "Health-Related Quality of Life of Severely Obese Children and Adolescents," *Journal of the American Medical Association*, vol. 289, no. 14, pp. 1813–1819, 2003, DOI 10.1001/jama.289.14.1813.

[3] Z. J. Ward, M. W. Long, S. C. Resch, C. M. Giles, A. L. Cradock, and S. L. Gortmaker, "Simulation of growth trajectories of childhood obesity into adulthood," *New England Journal of Medicine*, vol. 377, no. 22, pp. 2145–2153, 2017, DOI 10.1056/NEJMoa1703860.

[4] T. V. Kral and E. M. Rauh, "Eating behaviors of children in the context of their family environment," *Physiology and Behavior*, vol. 100, no. 5, pp. 567–573, 2010, DOI 10.1016/j.physbeh.2010.04.031.

[5] C. C. Yang and Y. L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010, DOI 10.3390/s100807772.

[6] J. Müller, A. M. Hoch, V. Zoller, and R. Oberhoffer, "Feasibility of physical activity assessment with wearable devices in children aged 4-10 Years-A Pilot study," *Frontiers in Pediatrics*, vol. 6, no. January, pp. 1–5, 2018, DOI 10.3389/fped.2018.00005.

[7] M. H. Pesch and J. C. Lumeng, "Methodological considerations for observational coding of eating and feeding behaviors in children and their families," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 14, no. 1, pp. 1–14, 2017, DOI 10.1186/s12966-017-0619-3.

[8] S. Bi, T. Wang, E. Davenport, R. Peterson, R. Halter, J. Sorber, and D. Kotz, "Toward a Wearable Sensor for Eating Detection," in *Proceedings of the ACM Workshop on Wearable Systems and Applications (WearSys)*. ACM Press, Jun. 2017, pp. 17–22, DOI 10.1145/3089351.3089355.

[9] S. Bi, T. Wang, N. Tobias, J. Nordrum, S. Wang, G. Halvorsen, S. Sen, R. Peterson, K. Odame, K. Caine, R. Halter, J. Sorber, and D. Kotz, "Auracle: Detecting Eating Episodes with an Ear-Mounted Sensor," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) (Ubicomp)*, vol. 2, no. 3, Sep. 2018, DOI 10.1145/3264902.

[10] A. Tada and H. Miura, "Association of mastication and factors affecting masticatory function with obesity in adults: A systematic review," *BMC Oral Health*, vol. 18, no. 1, pp. 1–8, 2018, DOI 10.1186/s12903-018-0525-3.

[11] R. M. Leech, A. Worsley, A. Timperio, and S. A. McNaughton, "Characterizing eating patterns: a comparison of eating occasion definitions," *The American Journal of Clinical Nutrition*, Oct. 2015, DOI 10.3945/ajcn.115.114660.

[12] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of Chewing Sounds for Dietary Monitoring," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2005, DOI 10.1007/11551201_4.

[13] S. Päßler, M. Wolff, and W.-J. Fischer, "Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food," *Physiological Measurement*, vol. 33, no. 6, pp. 1073–1093, Jun. 2012, DOI 10.1088/0967-3334/33/6/1073.

[14] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior." *Physiological measurement*, vol. 29, no. 5, pp. 525–541, May 2008, DOI 10.1088/0967-3334/29/5/001.

[15] M. Shuzo, S. Komori, T. Takashima, G. Lopez, S. Tatsuta, S. Yanagimoto, S. Warisawa, J.-J. Delaunay, and I. Yamada, "Wearable Eating Habit Sensing System Using Internal Body Sound," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 4, no. 1, pp. 158–166, 2010, DOI 10.1299/jamdsm.4.158.

[16] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G.-Z. Yang, "An Intelligent Food-Intake Monitoring System Using Wearable Sensors," in *Ninth International Conference on Wearable and Implantable Body Sensor Networks*. Hamlyn Centre, Imperial Coll. London, London, UK, May 2012, pp. 154–160, DOI 10.1109/bsn.2012.11.

[17] J. van den Boer, A. van der Lee, L. Zhou, V. Papapanagiotou, C. Diou, A. Delopoulos, and M. Mars, "The SPLENDID Eating Detection Sensor: Development and Feasibility Study," *JMIR mHealth and uHealth*, vol. 6, no. 9, p. e170, 2018, DOI 10.2196/mhealth.9781.

[18] V. Papapanagiotou, C. Diou, and A. Delopoulos, "Chewing detection from an in-ear microphone using convolutional neural networks," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2017, pp. 1258–1261, DOI 10.1109/EMBC.2017.8037060.

[19] R. Zhang and O. Amft, "Retrieval and Timing Performance of Chewing-Based Eating Event Detection in Wearable Sensors," *Sensors Journal*, 2020, DOI 10.3390/s20020557.

[20] E. S. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior." *IEEE Transactions on Bio-medical Engineering*, vol. 57, no. 3, pp. 626–633, Mar. 2010, DOI 10.1109/TBME.2009.2033037.

[21] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury, "BodyBeat: A Mobile System for Sensing Non-speech Body Sounds," in *Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2014, DOI 10.1145/2594368.2594386.

[22] T. Olubanjo and M. Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4384–4388, DOI 10.1109/icassp.2014.6854430.

[23] K. Yatani and K. N. Truong, "BodyScope: a wearable acoustic sensor for activity recognition," in *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp)*, ser. Ubicomp, Sep. 2012, pp. 341–350, DOI 10.1145/2370216.2370269.

[24] K. S. Chun, S. Bhattacharya, and E. Thomaz, "Detecting Eating Episodes by Tracking Jawbone Movements with a Non-Contact Wearable Sensor," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, Mar. 2018, DOI 10.1145/3191736.

[25] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen, "Image browsing, processing, and clustering for participatory sensing," in *Proceedings of the 4th workshop on Embedded networked sensors - EmNets '07*, 2007, pp. 13–17, DOI 10.1145/1278972.1278975.

[26] E. Thomaz, A. Parnami, I. Essa, and G. D. Abowd, "Feasibility of identifying eating moments from first-person images leveraging human computation," in *Proceedings of the 4th International SenseCam &*

Pervasive Imaging Conference on - SenseCam, ACM. ACM Press, 2013, pp. 26–33, DOI 10.1145/2526667.2526672.

[27] M. Sun, L. E. Burke, Z. H. Mao, Y. Chen, H. C. Chen, Y. Bai, Y. Li, C. Li, and W. Jia, "eButton: A Wearable Computer for Health Monitoring and Personal Assistance," in *Proceedings of the Annual Design Automation Conference*, 2014, DOI 10.1145/2593069.2596678.

[28] S. Sen, V. Subbaraju, A. Misra, R. K. Balan, and Y. Lee, "The case for smartwatch-based diet monitoring," in *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, Mar. 2015, pp. 585–590, DOI 10.1109/percomw.2015.7134103.

[29] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free-living by tracking wrist motion." *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1253–1260, Jul. 2014, DOI 10.1109/JBHI.2013.2282471.

[30] R. C. Klesges, T. J. Coates, G. Brown, J. Sturgeon-Tillisch, L. M. Moldenhauer-Klesges, B. Holzer, J. Woolfrey, and J. Vollmer., "Parental influences on children's eating behavior and relative weight." *Journal of applied behavior analysis*, vol. 16, no. 4, pp. 371–378, 1983, DOI 10.1901/jaba.1983.16-371.

[31] C. H. Llewellyn, C. H. Van Jaarsveld, D. Boniface, S. Carnell, and J. Wardle, "Eating rate is a heritable phenotype related to weight in children," *American Journal of Clinical Nutrition*, vol. 88, no. 6, pp. 1560–1566, 2008, DOI 10.3945/ajcn.2008.26175.

[32] N. Alshurafa, A. W. Lin, F. Zhu, R. Ghaffari, J. Hester, E. Delp, J. Rogers, and B. Spring, "Counting bites with bits: Expert workshop addressing calorie and macronutrient intake monitoring," *Journal of Medical Internet Research*, vol. 21, no. 12, pp. 1–16, 2019, DOI 10.2196/14904.

[33] T. Prioleau, E. Moore, and M. Ghovanloo, "Unobtrusive and Wearable Systems for Automatic Dietary Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2075–2089, Sep. 2017, DOI 10.1109/tbme.2016.2631246.

[34] T. Vu, F. Lin, N. Alshurafa, and W. Xu, "Wearable Food Intake Monitoring Technologies: A Comprehensive Review," *Computers*, vol. 6, no. 1, 2017, DOI 10.3390/computers6010004.

[35] O. Amft and G. Troster, "On-Body Sensing Solutions for Automatic Dietary Monitoring," *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 62–70, Apr. 2009, DOI 10.1109/mprv.2009.32.

[36] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "A Survey of Diet Monitoring Technology," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 57–65, Jan. 2017, DOI 10.1109/mprv.2017.1.

[37] S. Sen, V. Subbaraju, A. Misra, R. K. Balan, and Y. Lee, "Experiences in Building a Real-World Eating Recogniser," in *Proceedings of the 4th International on Workshop on Physical Analytics*, ser. WPA, 2017, pp. 7–12, DOI 10.1145/3092305.3092306.

[38] M. Mirtchouk, D. Lustig, A. Smith, I. Ching, M. Zheng, and S. Kleinberg, "Recognizing Eating from Body-Worn Sensors: Combining Free-living and Laboratory Data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 1, no. 3, pp. 85+, Sep. 2017, DOI 10.1145/3131894.

[39] R. Zhang and O. Amft, "Monitoring Chewing and Eating in Free-Living Using Smart Eyeglasses," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 23–32, Jan. 2018, DOI 10.1109/jbhi.2017.2698523.

[40] A. Bedri, R. Li, M. Haynes, R. P. Kosaraju, I. Grover, T. Prioleau, M. Y. Beh, M. Goel, T. Starner, and G. Abowd, "EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments," *Proc. ACM Interactive, Mobile and Wearable Ubiquitous Technology*, vol. 1, no. 3, Sep. 2017, DOI 10.1145/3130902.

[41] S. Zhang, Y. Zhao, D. T. Nguyen, R. Xu, S. Sen, J. Hester, and N. Alshurafa, "NeckSense: A Multi-Sensor Necklace for Detecting Eating Activities in Free-Living Conditions," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) (Ubicomp)*, vol. 4, no. 2, Jun. 2020, DOI 10.1145/3397313.

[42] E. A. Carroll, M. Czerwinski, A. Roseway, A. Kapoor, P. Johns, K. Rowan, and M. C. Schraefel, "Food and mood: Just-in-time support for emotional eating," in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 2013, pp. 252–257, DOI 10.1109/ACII.2013.48.

[43] T. Rahman, M. Czerwinski, R. Gilad-Bachrach, and P. Johns, "Predicting "about-To-eat" moments for just-in-Time eating intervention," in *DH 2016 - Proceedings of the 2016 Digital Health Conference*. New York, New York, USA: Association for Computing Machinery, Inc, Apr. 2016, pp. 141–150, DOI 10.1145/2896338.2896359.

[44] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos, "A novel chewing detection system based on PPG, audio and accelerometry," *IEEE Journal of Biomedical and Health Informatics*, 2016, DOI 10.1109/jbhi.2016.2625271.

[45] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: An audio analysis library for music information retrieval," *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013*, pp. 493–498, 2013, DOI 10.1145/2502081.2502229.