Wearable Computing and Healthcare

On-Body Sensing Solutions for Automatic Dietary Monitoring

Various on-body sensors can gather vital information about an individual’s food intake. Such data can both help weight-loss professionals personalize programs for clients and inform nutrition research on eating behaviors.

Maintaining a balance between ingested energy through food consumption and expended energy in daily life is key to a person’s long-term health. However, as the pandemic of overweight and obese individuals attests, this balance is challenging to maintain. According to World Health Organization estimates, more than one billion adults worldwide were overweight and 400 million were obese in 2005 (www.who.int/topics/obesity/en/). By 2015, WHO predicts the number of obese people alone will increase to more than 700 million.

To support active weight control, various public and private organizations have established weight and diet management programs. Such programs coach individuals to improve eating behaviors using daily or weekly status feedback, meal suggestions, and behavior recommendations. However, as Rena Wing and Suzanne Phelan found, only 20 percent of the people who achieve at least a 10 percent reduction in body weight can maintain that new weight for one year. They therefore concluded that, to increase coaching program success, participants need two to five years of continuous support. In addition, further eating behavior research is needed to advance our understanding of human drive and restraint regarding food intake. The main limitation of current programs aimed at achieving this is a practical one: Participants must complete detailed self-reports on their eating behavior, while also maintaining their changed lifestyles and eating behavior on a day-to-day basis.

Along with a personal profile, self-reports are currently the sole source of information for adapting and personalizing feedback and recommendations for coaching program participants. Unfortunately, self-reports have a high bias and are hard to maintain. A key shortcoming is that individual respondents can differ dramatically in their motivation to complete questionnaires, their awareness of food intake (particularly with snacks), their literacy level, and their memory and perception capabilities.

Dale Schoeller, for example, found that self-reporting estimations ranged from 50 percent under to 50 percent over the actual intake amount.

To address these issues, automatic dietary monitoring (ADM) aims to replace manual eating-behavior reporting with a sensor-based estimation. Here, we discuss requirements and options for on-body sensing of eating behavior, and demonstrate that such sensor information can resemble some self-report information. These initial ADM research prototypes aren’t yet comfortable enough for continuous

Oliver Amft
TU Eindhoven

Gerhard Tröster
ETH Zurich
Diet Monitoring Approaches

Classic dietary monitoring techniques require users to manually record their eating behavior. Among these assessments, self-reports are intended to capture every food intake, on a day-to-day basis, as required by weight and diet management programs. However, low adherence and accuracy restrict the reports’ validity and thus the benefit of coaching programs that use them.

Researchers have made multiple attempts to simplify tedious and error-prone logging. However, studies have confirmed that replacing paper-based reports with manually operated electronic devices doesn’t reduce reporting errors.

Manual Methods

There are several alternate manual methods for capturing eating behavior information. Jennifer Mankoff and her colleagues scanned shopping receipts to simplify diet monitoring. MyFoodPhone Nutrition, Inc. (www.myfoodphone.com) introduced a commercial service that assesses food intake based on users’ mobile-phone pictures. Katie Siek and her colleagues used bar codes and voice recordings to replace self-report questionnaires.

Automated Solutions

With manual dietary monitoring, participants are asked to record their own eating behavior. In contrast, automatic dietary monitoring aims to estimate eating behavior without the person’s active participation. We can categorize these automated techniques according to their sensing approach: ambient-embedded, on-body, or implantable.

Researchers have developed a few pioneering solutions that use ambient-embedded sensors. Keng-Hao Chang and his colleagues developed a dining table that detected the weight of foods and identified food bowls from radio-frequency identification tags. Jiang Gao and colleagues used surveillance video to identify arm movements to the mouth. In their general evaluation of RFID for home monitoring, Donald Patterson and his colleagues estimated morning activities, including breakfast consumption timing.

Implantable solutions, such as in-oral sensing, could provide more precise information on the eating process. However, this solution is technically challenging and alters oral sensation. Hence, it appears infeasible for long-term diet monitoring.

REFERENCES

• food amount, and
• energy content (calories).

The AMD systems must also meet operational requirements, prove robust, and offer comfort suitable for long-term use.

Challenges for ADM
For self-reports and ADM solutions, the challenge is to capture both the diversity of consumed foods and the variability in personal eating behaviors. For example, energy intake is most accurately determined by reporting the consumed food’s calories. However, even with direct calorie reporting, energy estimation requires additional information, including the amount of consumed food and whether the person has altered it (such as by adding a dressing or sauce). Furthermore, calorie reporting is often complex and infeasible for homemade meals.

People have preferences about their food choices and categories, and their meal schedules. ADM solutions can integrate these preferences as prior information to estimate eating behavior. Still, a person’s actual eating behavior is influenced by varying environmental and psychological aspects, including constraints in food availability, social interaction during meals, and emotions.

A particular challenge for ADM solutions is to robustly recognize eating behavior from the sensor data. No single sensor—independent of its location and recorded physiological or activity information—can capture all dimensions of eating behavior. The restrictions of initial ADM approaches reflect this challenge. Typically, the solutions emphasize particular dimensions of eating behavior, such as recording consumed food amounts using a weight scale, while restricting location to the weighting-enabled table. Moreover, solutions that rely exclusively on ambient-embedded sensors increase the challenge of robustly assigning measurements to one person. Although these works represent relevant advancements toward ADM, a multimodal sensing approach will better support monitoring of several eating behavior dimensions.

Benefits of On-Body Sensing
Monitoring eating behavior in a continuous and location-independent way is a vital ADM system property, as modern lifestyles imply location changes for both work and leisure purposes. Consequently, people consume food in various locations and in transit. Solutions that depend on a particular environment—such as the home—would miss snacks, let alone entire business lunches. Such partial coverage severely limits the effect of behavior coaching and could produce misleading recommendations.

Coaching and research-oriented behavioral understanding require continuous diet monitoring that covers all situations. On-body sensors can continuously monitor eating behavior, independent of dedicated sensor-enabled environments. Also, in contrast to ambient-embedded sensors, on-body sensors directly associate recorded information with the wearer.

Evaluation of On-Body Sensing Solutions
We analyzed on-body sensing approaches and modalities to evaluate the benefits for ADM. As Figure 1 shows, our analysis covered activities related to eating and physiological responses to food consumption. We analyzed which eating behavior dimensions a particular solution helps to estimate, as well as its limitations. We also assessed how comfortable it is to wear.
Assessment Criteria
As we noted earlier, estimating energy intake requires the food’s category and amount, combined with a more complex inference. We therefore didn’t include energy intake in this investigation. Table 1 summarizes our evaluation results for all sensing solutions on eating behavior dimensions, limitations, and comfort.

We selected and analyzed three basic eating activities: intake gestures, chewing, and swallowing. These activities represent a temporal description of food intake and help identify intake cycles. We developed sensing prototypes for these activities and analyzed the effectiveness of these solutions for predicting food category and amount in user studies. To obtain individual performance estimates, we used a Naïve Bayes classifier preceded by linear discriminate feature extraction. Finally, to ensure the results’ robustness, we deployed a five-fold cross-validation.

Intake Gestures
Most food intake requires upper body movements (arms and trunk). We distinguished these movements into coarse food and beverage preparation—such as unpacking, cooking, and plate loading—and actual food intake. The latter includes motions to fine-cut and maneuver the food piece to the mouth. In the intake phase, people use tools, such as forks and knives. We focused our recognition approach on intentional arm movements for the intake, which we refer to as “intake gestures.” Because these intake gestures reflect intake types (eating or drinking) and food category (based on tools used), they provide timing and food category information.

We record intake gestures using inertial sensors at the participant’s wrists and upper back. As Figure 2a shows, we derived a comfortable recording setup by integrating commercial motion sensors (www.xsens.com) in a jacket. The sensing units contain 3D acceleration, gyroscope, and compass sensors.

To evaluate how the sensors help to discriminate between different gestures, we studied four students eating foods in four different gesture categories:

- eating lasagna with a fork and knife,
- drinking from a glass,
- eating soup with a spoon, and
- eating bread using only one hand.

The students ate and drank in random order, without particular movement instructions. During recording breaks, they performed other activities—such as reading a newspaper and making a phone call—to promote natural movement variability. In total, we recorded 1,020 intake gestures over 4.68 hours.

Using the classification procedure, we obtained 94 percent accuracy overall. Figure 2b shows the results for individual gesture categories. We used only temporal features from arm acceleration sensors. We observed that we can model intake gestures’ temporal structure by computing each gesture instance’s features in four sections. Without these temporal features, we achieved similar classification results, but had to use all motion sensor modalities and hidden Markov models.

Although the motion sensor jacket was a useful research prototype, we plan to replace it with less complex
sensors. The classification using only acceleration shows that we can reduce the number of sensors. That said, our study wearers reported that the jacket was comfortable for sitting activities.

**Chewing**

One option for recording chewing strokes (the jaw opening and closing) is to monitor masseter and temporalis muscle activation using surface electromyography (EMG). Because muscles are located in exposed facial regions, sensing jaw movement—which is highly variable during chewing and other motions, such as speaking—might require attaching a sensor in exposed facial regions.

To avoid compromising privacy, we found a feasible alternative. Chewing generates sound emissions that conduct through mandible, skull, and body tissue. So, we recorded chewing sounds using an ear-attached microphone. Based on an acoustic profile during chewing, we classified foods\(^6\) and analyzed different microphones and ear-device cases. Figure 3a shows one device, in which we embedded a miniature microphone in a standard headphone case. In another construction, we used an ear-pad case. With the latter setup, we studied how users perceived the ear occlusion. Smaller pads reduced ear occlusion and increased user comfort, but also reduced the signal-to-noise ratio. Users found the headphone device convenient; this was especially true of those who were used to wearing similar models with music players.

We studied the scalability of this approach to classify various foods. To do this, we asked three male students with natural dentition to eat 19 standard foods as they normally would. In several sessions, we recorded chewing using a low-occlusion ear-pad device. In this setup, the wearer could understand office-room conversation within two meters. In all, we obtained approximately 12,000 chewing strokes in five hours of data. For classification of all foods, we obtained accuracy of 80 percent. For the headphone case, this high accuracy dropped by five to 10 percent, depending on ambient noise. As features, we used spectral energy bands and cepstral and linear predictive coefficients.\(^6\) We selected these features based on robust results obtained with earlier recordings.

Figure 3b offers a quick overview of classifier performance for all foods. The color-coding shows classifier confusions (the yellowish colors that fall outside the main diagonal) of various acoustic groups among foods. Sound patterns are primarily controlled by food texture—and, thus, lettuce is partly confounded with carrots and apples.

In this evaluation, food texture was our main selection criteria. The set includes similar textures, such as lettuce and apples, and covers a broad variety of materials and preparation styles, such as for cooked meat. Because our result demonstrates texture-based discrimination capabilities, we further deploy chewing sound recognition for nutritional-relevant food groups from the food pyramid. For example, we can group fruits and vegetables based on a similar “wet-crisp” texture and recognize this group in continuous sound data.\(^6\)

**Swallowing**

Swallowing, which happens unconsciously throughout the day, occurs with increased frequencies during food intake.\(^7\) Specifically, after we chew food and convert it into a bolus, our tongue movements initiate a reflex of throat muscles that propel the bolus through the throat into the esophagus.

Most swallowing studies analyze ab-
normal swallowing in laboratory settings. Because tongue and esophageal movements are challenging to monitor with on-body sensors, we focused on the swallowing reflex using sensors at the throat. To investigate different sensing modalities, we developed a set of collars.

As Figure 4a shows, in one collar system, we monitored textile elongation to detect skin movement during swallowing. Such elongations occur mainly for male subjects, since females have a less prominent Adam’s apple. Moreover, the strain-sensing collar required accurate positioning, and signals were impaired when the neck and collar moved.

In a second solution, we combined surface EMG and a stethoscope-like microphone to monitor both throat movement and swallowing sounds (see Figure 4b). While EMG is impaired by other throat-muscle activations, the swallowing sound pattern is influenced by food viscosity. We combined both modalities to determine the amount of swallowed food.

We used these sensors with five students eating foods and drinking water as they normally would. Over several sessions, we analyzed a total of 4,850 hours of data and 868 swallows. We discriminated between two types of swallowing: low volume (such as 5 milliliters of water, a spoonful of yogurt, and 2 cubic centimeter pieces of bread) and large volume (15 ml water) with an overall accuracy of 73 percent. As with chewing sound classification, swallowing volume discrimination required a spectral feature set.

As expected, users found both collars uncomfortable for long-term monitoring. Our current work aims to replace the collar prototypes with more convenient systems, such as embedding sensors in a shirt collar.

Further On-Body Sensing Options
We analyzed whether other sensing solutions could provide eating behavior information. Our goal was to review activities and physiological responses closely related to food intake and summarize potential benefits for ADM.

Gastric Activity
Swallowed food arrives at the stomach after roughly 15 minutes. It’s subsequently decomposed by stomach muscle contractions. Further digestion in the gastrointestinal tract incurs time delays in the range of hours with respect to the originating intake and thus is far less deterministic.

There are few on-body sensing options for late digestion stages. Researchers have captured stomach muscles’ electric and magnetic fields using laboratory setups, such as electrogastrography (EGG). However, EGG hasn’t reached broad clinical acceptance. A stethoscope can assess abdominal sounds from food movement in intestines. Although bowel sounds are typically loudest after fasting, researchers recently confirmed a relation to intake for laboratory settings. All such measurements are perturbed by heart and respiration activity, as well as by body movement.

Thermic Effect of Food Intake
The thermic effect of food intake (TEF) is a thermogenesis in response to intake above resting metabolic rate. Although TEF is the smallest component in human energy expenditure, researchers studied its relation to intake restraint and obesity.

Optimal TEF assessment requires a respiratory chamber to measure changes in resting metabolic rate before and after intake. TEF starts immediately after food reaches the stomach and peaks after roughly 60 minutes. For unrestrained eating in people of normal weight, skin temperature above the liver increased between 0.8 and 1.5K. TEF depends on regularity of intake and is lower for irregular intake.

Body Weight
Food intake is associated with an immediate gain in body weight. If weight is monitored, we can determine intake timing and food amount. Typical meals range from 50 grams for light meals to 500 grams or more for multiple-course menus. Snacks can weigh just a few grams, but still contribute an important share to daily intake, as they often include high-calorie foods or sweets.

In contrast to classic body weighting (such as weekly measurements), intake-related weight changes require continuous day-long weighting. Load sensors in shoes would ideally serve this purpose. Compared to a scale, shoe-based weighting requires a low mechanical profile, high torsion flexibility, and low system weight. Also, the system must measure weight from foot force distribution in the (sometimes brief) moments when the user is standing still. These requirements are not easily met. Classic load cells don’t fulfill the mechanical constraints. Pressure-sensing
### TABLE 1
Assessing on-body sensing solutions for automatic dietary monitoring (ADM).

<table>
<thead>
<tr>
<th>Sensing solution</th>
<th>Dimensions of eating behavior</th>
<th>Modalities and comfort for everyday use</th>
</tr>
</thead>
</table>
| **Intake gestures** | **Timing**: continuous recognition of four gesture types yielded R: 79 percent, P: 73 percent.  
**Food type**: movement related to food category; recognizes four types at C: 94 percent.  
**Food amount**: open  
**Limit**: errors occur for arbitrary arm movements to the head and unusually long gestures. | **Modalities**: inertial sensors at lower arms and upper back.  
**Comfort**: conveniently integrated into smart clothing or accessories, such as a watch or bracelet. |
| **Chewing** | **Timing**: continuous recognition for two food categories yielded R: 93 percent, P: 52 percent.  
**Food type**: ear-pad device recognizes 19 foods using chewing strokes. C: 80 percent.  
**Food amount**: open  
**Limit**: perturbed by ambient noise and low ear occlusion. | **Modalities**: ear-pad microphone, similar to ear-attached hearing-aid devices.  
**Comfort**: depends on ear occlusion; convenient for headphone case. |
| **Swallowing** | **Timing**: continuous recognition of four bolus types yielded R: 65 percent, P: 31 percent.  
**Food type**: open  
**Food amount**: bolus volume; recognizes low vs. high volume yielded C: 73 percent.  
**Limit**: individual modalities impaired by head and neck movements, chewing, and speaking. | **Modalities**: surface electromyography, stethoscope microphone, or similar acoustic transducer; skin movement at the throat (as in our studies); throat impedance or capacitive sensing.  
**Comfort**: large size sensor-collars are uncomfortable; improvements expected with collar-shirt implementations. |
| **Gastric activity** | **Timing**: stomach activity increases roughly 15 minutes after intake; duration dependencies unclear.  
**Food type & amount**: relations unclear.  
**Limit**: approaches require strict laboratory settings; infeasible for nonstationary monitoring. | **Modalities**: electrogastrography, impedance gastrography, and bowel sounds.  
**Comfort**: electrodes/sensors must be tightly attached to chest or belly. |
| **Thermic effect** | **Timing**: temperature increase of 0.8 to 1.5 Kelvin roughly 60 minutes after intake; duration dependencies unclear.  
**Food type & amount**: relations unclear.  
**Limit**: temperature depends on regularity of food intake. Unrestricted physical activity and ambient temperature alter this relationship. | **Modalities**: temperature sensor.  
**Comfort**: must be attached to skin in proximity of the liver. |
| **Body weight** | **Timing**: body weight increases immediately; intake duration isn’t assessable.  
**Food type**: n/a  
**Food amount**: required weight monitoring resolution is <50 grams for meals and 5 grams for snacks.  
**Limit**: shoe-based weight measurement requires users to stand still and is impaired by uneven floor surfaces. | **Modalities**: shoe-embedded weight or force sensor array (unsolved); current in-shoe force sensors don’t provide appropriate resolution.  
**Comfort**: related to shoe torsion flexibility and weight. |
| **Cardiac responses** | **Timing**: heart rate increases roughly 30 minutes after intake for up to 3 hours (laboratory).  
**Food type & amount**: relation to heart rate is unclear; blood pressure is influenced by salt and sugar.  
**Limit**: relationships altered by physical activity, fasting time, and time of day. Measurements perturbed by physical activity. | **Modalities**: for heart rate, electrocardiogram chest strap or close-fitting shirt; for blood pressure, cuff-based or cuffless monitor. Research on cuffless approaches is ongoing.  
**Comfort**: cuff-based monitor impractical for long-term use. |
| **Body composition** | **Timing**: body impedance altered roughly 30 minutes after intake in clinical settings; duration unknown.  
**Food type & amount**: relations unclear.  
**Limit**: measurements perturbed by body movement. | **Modalities**: body impedance using electrodes.  
**Comfort**: hand-to-foot electrodes are potentially inconvenient. |

*R = recall, P = precision, and C = classification accuracy
arrays struggle to meet weight resolution requirements. Capacitive in-shoe gait measurement systems have an error of 2.7 percent,\textsuperscript{13} corresponding to 1,890 grams for a 70-kilogram person. We studied arrays of force-sensitive resistors and observed even larger errors due to signal noise and shoe torsion. At this point, a continuous wearable measurement of body weight remains unsolved.

Cardiac Responses
After meal intake, blood is redistributed to the stomach and lower gastrointestinal tract, which increases heart rate 30 minutes after intake.\textsuperscript{14} Blood pressure is dependent on food composition, especially on salt and sugar. Classic blood pressure measurements require cuff-based solutions and are inconvenient for daily use. However, ongoing research is investigating novel cuffless approaches, such as those based on pulse arrival time. Cardiac responses depend on various aspects, including physical activity, body posture, fasting time, and time of day.

Body Composition
Food intake immediately modifies body composition. In a laboratory setting, we measure body impedance between the hand and foot; studies show that composition is altered 30 minutes after intake.\textsuperscript{15} The effect depends on both gender and food type, and further investigations are needed to study composition assessment validity. In any case, movement artifacts make the effect impractical for ADM systems.

**Intake Cycle Modeling**
Intake gestures, chewing, and swallowing represent a temporal description of food intake. As Figure 5a shows, we selected these solutions to construct a hierarchical recognition procedure to identify intake cycles. In our approach, an intake cycle stretches from an intake gesture (taking a bite of food) until the bite is completely swallowed. We deployed individual detectors to recognize activity events from each sensing solution.

Figure 5b illustrates two event sequences—the intake cycles for drinking and for eating. To recognize intake cycles from activity events, we implemented a probabilistic context-free grammar parser.\textsuperscript{16} The PCFG estimates the fit of event sequences to an intake grammar. We derived grammars for particular food categories, such as drinking and eating fruits. PCFGs let us model recursive event structures, such as the recursion of chewing and swallowing events for eating an apple (Figure 5b).

Our approach provides a number of benefits for estimating eating behavior:

- The temporal fusion of individual food category estimations from intake gestures and chewing lets us recognize more diverse categories.
- The fusion complements individual sensing solutions’ estimation errors.
- At the event level, hierarchical recognition allows simplified synchronization of sensing solutions with different sampling rates.

Because ADM aims to replace manual monitoring for weight and diet coaching, we can use the manual method’s eating behavior information requirements and benchmarks for ADM solutions. In our evaluations, we observed that recognizing intake activities from on-body sensors provides information on intake timing, food category, and amount. Moreover, by using on-body sensors, information is obtained continuously, independent from particular locations. Nevertheless, many current on-body sensing solutions have limitations regarding data artifacts and wearer comfort.

Although combining selected solutions in a hierarchical recognition can compensate for individual sensors’ estimation errors, it refines estimations for food categories only. In comparison to self-reports—which could capture in-
formation on exact food type—this is a limitation. Similar restrictions apply for food amount (and hence energy intake estimation). While energy intake is important, food also provides essential nutrients, and individual nutrient requirements can vary widely. Moreover, eating disorders—such as binge eating—indicate that eating is tightly coupled to momentary psychological state and emotions. Self-reports could ask specific questions to capture this day-to-day variation. However, if we consider self-reporting’s practical issues and biases, even the categorical information obtained with ADM is highly beneficial. We expect that initially deployed systems will track a few food categories, such as fruits and vegetables, related to particular nutritional recommendations. Our studies showed high recognition performances for identifying these categories.

Among all selected sensing solutions, the least comfortable are the swallowing solutions. In our on-going research, we plan to replace the current collar prototypes with more convenient systems. In addition to the diet coaching domain, we plan to deploy ADM-sensing solutions in basic research to advance the understanding of eating behavior. Finally, we plan to combine on-body and ambient sensing solutions to leverage the advantages of both approaches.

ACKNOWLEDGMENTS

We are grateful to all project participants for committing time to our various studies. The Swiss State Secretariat for Education and Research and the European Union’s MyHeart (IST-2002-507816) project partially supported this work.

REFERENCES


