Every year, significant resources are invested to improve medical treatments, develop pharmaceutical drugs, and mitigate the effects of various diseases. However, less emphasis is placed on preventative healthcare. By ensuring that individuals exercise regularly and eat a balanced diet, we can reduce their chances of various health risks such as cancer, diabetes, and heart disease. Although public health measures have sought to encourage individuals to lead healthy lifestyles, quantifying diet and exercise has generally remained an unaddressed challenge. Only recently has accurate evaluation of physical activity become a reality, because numerous noninvasive activity monitoring systems—such as FitBit (www.fitbit.com), MisFit (www.misfitwearables.com), and Jawbone (http://jawbone.com)—have entered the market. However, resources for evaluating diet are less prevalent, as most techniques outside of the academic literature focus on manual record keeping. These methods are generally inaccurate and inefficient, and they present a burden to individuals who simply fail to adhere to the record-keeping procedure for extended periods of time.

This article evaluates the potential of various approaches to dietary monitoring with respect to convenience, accuracy, and applicability to real-world environments. We emphasize the application of technology and sensor-based solutions to the health-monitoring domain, and we evaluate various form factors to provide a comprehensive survey of the prior art in the field. Although we can’t cover every approach, we review most major techniques with an emphasis on works that can be applied to real-world environments rather than constrained to laboratory settings.

Figure 1 illustrates several of these approaches. The aims of these approaches vary from suggestions on eating pace, identification of food volume, and even detection of food composition. Prior research has shown that macronutrient compositions of foods can yield different health outcomes.

Through an online survey we evaluated individual perceptions of some of the most promising techniques for nutrition monitoring in nonclinical settings, including histograms of interest in various devices and comments from individuals with respect to the proposed ideas.

Manual Record Keeping

Wearable nutrition monitoring devices are becoming a subject of interest in the research community. However, the most pervasive techniques today are still manual record-keeping approaches. Although these techniques are simple and affordable, they’re associated with user burden and poor accuracy. Furthermore, although manual techniques are often conducted without electronic technology, several recently proposed applications, such as MyFitnessPal and SparkPeople, provide data logging capabilities and other resources for weight management. For
brevity, we limit our discussion here to several major techniques.

**Multipass 24-Hour Recall**
A simple, pervasive method of monitoring diet is the multipass 24-hour dietary recall method, which is based on the data that patients provide at the end of a randomly selected day. Each individual gives an oral or written report including the amount and type of food eaten during the day, which is used to calculate total food intake. This approach measures food intake in a reasonably quantitative manner but with significant error because people do not recall the exact amount of food they have eaten. 11 Experimental data shows that food intake is usually reported with error, and measurement variance also depends on the patient’s experience with the system. 12

**Food Records**
Food records are generally not impacted by the accuracy of a subject’s memory. They typically require individuals to note their eating habits during or immediately after a meal. Concerns include patient compliance and the difficulty that untrained individuals face when accurately assessing portion size. For example, caloric density varies based on cooking method and other factors that are not necessarily visually apparent. 13 Moreover, manually recording each meal can be tedious, and many individuals will be unwilling to complete food records for extended periods of time.

**Food-Frequency Questionnaire**
A third method for manually assessing dietary intake is to use a food frequency questionnaire (FFQ), in which individuals specify their rate of consumption for various food items over the long term. Nutritional intake can subsequently be assessed by summing various food types provided within the list. 13 This technique is inexpensive to administer and insensitive to recent changes in diet. Furthermore, it is less time-consuming than food records because it is not intended to be completed on a daily basis. However, FFQs are typically less accurate than other techniques. This is often a result of several factors, including incomplete lists of food, poor user compliance, and frequency and serving size errors. 14

**Review of Manual Recording**
Manual record-keeping approaches are simple, and don’t require users training for custom hardware operation. The initial cost is also relatively low, because they don’t require the purchase of any expensive devices.

On the other hand, manual record keeping can be tedious, and long-term compliance rates are low. In addition, individuals often do not accurately remember what they have eaten or might deliberately misreport their eating habits.

**Acoustic Methods**
Audio-based methods are among the most popular approaches for monitoring eating habits in the academic literature. They generally involve placing microphones near the throat to record chewing and swallowing sounds.

Andrea Santamato and his colleagues analyzed the acoustic properties of swallow sounds for semisolids, semiliquids, solids, and liquids, and found that the maximum frequencies, which ranged from 2,281.3–4,244.0 Hz, were associated with liquids. 15 Thus, effective characterization of swallow sounds might require sample rates of at least 8.4 kHz based on Shannon-Nyquist sampling theory. Very high-speed data acquisition and transmission is also associated with significant power overhead. 16

**Throat Microphone**
In 2010, Edward Sazonov and his colleagues proposed a technique for predicting food intake based on audio signals acquired from a throat microphone with a frequency range of 20–8,000 Hz. 17 The dataset used in their experimental evaluation was substantial,
Figure 2. An illustration of Koji Yatani and Khai Truong’s proposal\textsuperscript{18} for detecting swallow sounds using a headset placed with the microphone pointing towards the neck. This arrangement lets the system detect a variety of activities while rejecting external sounds. It is one example of an acoustic method for diet monitoring.

with a total of 9,966 swallows from 20 subjects. The authors placed an IASUS throat microphone over the subject’s laryngopharynx to minimize the distance between the microphone and the source of the swallow sound compared to microphones placed in the ear. Using techniques based on the Mel-scale Fourier spectrum, wavelet-packet decomposition, and support vector machines (SVMs), they identified swallow events with 84.7 percent weighted accuracy despite the presence of various artifacts, including respiration, speech, and head movements. Although this classification accuracy represents just two classes (swallowing or other), we used three foods with varying textures during experimentation: pizza, apples, and sandwiches.

**Headset Microphone**

In 2011, Koji Yatani and Khai Truong presented BodyScope, a Bluetooth headset with an embedded microphone for recording throat noises.\textsuperscript{18} The device rests on the lower neck, with an earpiece pointed inward toward the skin, as Figure 2 illustrates. The microphone is attached to a stethoscope chestpiece to filter external sounds that do not originate from the neck. The system analyzes eating, drinking, speaking, coughing, and eight other activities in a laboratory environment with 78.5 percent classification accuracy.

In a real-time evaluation, it recognized a subset of four activities with 71.5 percent accuracy. Classification was achieved using SVMs, a radial basis function (RBF) kernel, and statistical features obtained from spectrogram representations of the audio signals. Results for leave-one-sample-out cross-validation are promising, given the large number of classes of data. However, leave-one-subject-out cross-validations significantly reduce classification accuracy, suggesting the need for individual calibration.

**Integrated Ear-Canal Microphone**

Several works have analyzed eating sounds using in-ear microphones. In 2012, Sebastian Päßler and his colleagues presented an audio-based approach for audio signals analysis to detect swallowing sounds.\textsuperscript{19} The hardware device consists of a hearing-aid package with two integrated microphones acquiring data at approximately 11 kHz, along with associated amplification and filtering. Classification results were promising: the system detected eight classes of food with an overall accuracy of 79 percent using hidden Markov models to analyze both chewing and swallowing sounds. A significant novelty of this system was its use of a reference microphone placed outside of the subject’s ear, in addition to the in-ear microphone. The authors used the ratio of signal energy from both microphones to distinguish external environmental sounds from those related to eating.

**Review of Acoustic Methods**

Advantages of audio-based techniques include their ability to identify specific foods, rather than eating events, with relatively high accuracy. They are also versatile, able to detect many activities by analyzing audio signals. Thus, these systems are useful for a variety of applications, such as exercise monitoring.

However, many individuals would object to wearing a headset or custom earpiece throughout the day. Battery life is also an issue, with audio signal processing requiring higher sample rates than inertial sensing.

**Gesture Recognition**

Individuals generally make distinct and recognizable gestures with their arms as they eat. Examples include picking up a sandwich, raising a glass of water to their mouths, scooping up food with a fork, or cutting a steak with a knife. Hardware devices mounted on the subject’s arm or wrist with embedded inertial sensors such as accelerometers and gyroscopes can recognize these gestures and infer eating activity. Whereas audio signals must be acquired using sample rates of up to 8 kHz, activity-recognition techniques use much lower sample rates, such as 100 Hz or below.\textsuperscript{20} The power savings of this approach can be significant, because transparency in wearables necessitates long intervals between recharges, as well as miniature designs with small batteries.

**Custom Sensor Nodes**

Oliver Amft and Gerhard Tröster explored different methods for monitoring food intake, including gesture recognition.\textsuperscript{20} Their experiments used sensor modules including triaxial gyroscopes, accelerometers, and compass sensors to detect four food-related gestures: cutting lasagna with a fork and knife, drinking from a glass, eating soup with a spoon, and eating sliced bread with one hand. The sensors were attached to a jacket in the lower and upper arm, as Figure 3 shows. To present a realistic experimental procedure, subjects periodically performed other motions such as reading a newspaper and making telephone calls. The system detected most activities, with the exception of eating...
bread, with high accuracy: 94 percent classification accuracy based on 1,020 recorded intake gestures.

Despite the high results, the multitude of sensors could be somewhat unwieldy for real-world situations; more practical realizations of these schemes would probably have sensors in a single location. For example, Yujie Dong and his colleagues proposed a bite-counter device shown to have a sensitivity of 91 percent by analyzing the rolling movement of the wrist to detect biting behavior.\(^{21}\)

**Smartwatch Platform**

Several works have recently proposed using smartwatch-based inertial sensors for gesture recognition to infer eating behavior. Sougata Sen and his colleagues use a smartwatch’s accelerometer and gyroscope to determine eating episodes with 97 percent accuracy, as well as the mode of eating (spoon, hands, or chopsticks) with 85.5 percent accuracy.\(^{22}\) They mention that in the future, the smartwatch’s camera could be automatically triggered to allow identification of the specific food based on image processing.

Other works have proposed techniques for identifying gestures such as those associated with medication adherence\(^{23}\) and lifestyle monitoring (such as jogging, walking, smoking, speaking).\(^{24}\) In the coming years, many activity-monitoring and gesture-recognition schemes will likely use smartwatches rather than sensor nodes. Moreover, the multitude of sensors in modern smartwatches provide an opportunity for aggregating data from cameras, inertial sensors, and microphones to increase accuracy.

**Review of Gesture Recognition**

Advantages of gesture recognition approaches include comfort. A lightweight sensor on the arm might be sufficient for detecting many eating gestures, making it practical for day-to-day use. In addition, activity-recognition techniques do not require the high sample rates that audio-based solutions require, so can potentially be powered for days using a small coin cell battery.

The main disadvantage of these approaches is their scope. Gesture recognition can provide only general information about eating habits with respect to meal consumption and timing. Furthermore, it’s unclear whether a single sensor node can accurately detect snacking with one hand or eating “on the go.”

**Instrumented Objects and Places**

In this section, we provide an overview of nonwearable techniques for diet monitoring, such as sensor-equipped dining accessories.

**Smart Utensils**

Smart utensils are another inertial-sensing approach for nutrition monitoring (see Figure 4). Azusa Kadomura and his colleagues propose a sensing fork with an embedded accelerometer, electrode, and color sensor.\(^{25}\) The electrodes detect eating behavior such as biting and poking based on grip, while the color sensor detects the food color, from which it can infer when the user switches from one food type to another. These sensors can collectively determine the utensil state: at rest, being held, and poking. A mobile application accompanying the fork aims to improve children’s eating habits with feedback on meal timing and composition.

The commercially available HapiFork offers a simpler approach.\(^{26}\) HapiFork incorporates inertial sensors into a smart-fork package. It detects movement of the hand to the mouth and vibrates when the eating pace is too fast. The fork can also measure how long it takes to finish a meal and the approximate serving size.

**Smart Tablecloth**

Bo Zhou and his colleagues presented a smart tablecloth in 2015.\(^{27}\) The system detects eating behavior on solid surfaces (such as tables) based on

---

**Figure 3.** An illustration of Oliver Amft and Gerhard Tröster’s proposed approach,\(^{20}\) which uses sensor nodes to identify gestures associated with eating behavior. These nodes, which include accelerometers, gyroscopes, and compass sensors, were used to detect eating gestures with a variety of utensils. This approach is representative of using gesture recognition to monitor diet.

**Figure 4.** Smart utensils with triaxial accelerometers and gyroscopes are one example of instrumented objects and places that can be used for diet assessment without requiring body-worn sensors.
because these accessories aren’t worn. However, these techniques can be of limited use in unconventional eating scenarios, such as snacking while performing other activities.

**Camera-Based Techniques**

Image processing and computer vision algorithms can be used to recognize eating and even predict the nutritional composition of foods. Here, we provide an overview of visual diet assessment techniques.

**eButton**

The eButton, shown in Figure 5, is a chest-mounted button with an embedded camera. The button is attached to a shirt using a pin or pair of disk magnets, and contains an ARM Cortex processor, two wide-angle cameras, a UV sensor for distinguishing between indoor and outdoor environments, inertial sensors, proximity sensors, a barometer, and a GPS.

The eButton takes photos at a preset rate, thereby recording the entire eating process. It uses image processing techniques to detect the dining accessories (such as a plate or bowl). Subsequently, the system identifies the food items based on color, texture, and other heuristics. Using this information and additional techniques, it calculates volume and calorie count for each food based on a public domain database. Researchers evaluated 100 foods with a calorie estimation error of approximately 30 percent for 85 percent of the foods, which were regularly shaped. However, the system didn’t detect irregularly shaped food with high accuracy.

**PlateMate**

With PlateMate, users take photos of their meals to receive information about food composition and nutrition. The system uses the Amazon Mechanical Turk crowdsourcing platform to offload the labor of identifying and labeling foods to the community. The authors used a food database to convert the community descriptions of portion and composition to a calorie estimate. The average error rate was 33.2 percent, which significantly outperforms self-reports.

**Review of Camera-Based Techniques**

Camera-based techniques can often identify specific foods, rather than reporting meal events, which can provide a more complete nutritional profile beyond volume or swallow count. However, the 30 percent error for most foods is high, considering an individual’s daily caloric intake might not vary by such a large margin.

**Chewing and Swallowing Motion**

Several works attempt to characterize eating habits based on skin motion during chewing and swallowing events. The primary challenges here relate to comfort, because an inertial sensor is typically required to be contact with the skin during the meal.

**Capacitive Sensor**

In 2010, Jingyuan Cheng and his colleagues proposed a capacitive-sensing neckband. Using conductive textile-based electrodes, they were able to classify a broad range of activities, including chewing, swallowing, speaking, and various head motions. Their system uses a feature-similarity search (FSS) algorithm and time-domain features such as signal mean, variance, minimum, and maximum. The hardware consisted of a neckband with various electrodes, a multistage amplifier, a low-pass filter, and a 24-bit analog-digital converter (ADC). Experimental results suggested that capacitive sensing can be a viable method to detect chewing and swallowing, and results are comparable to prior approaches involving audio and electromyography.

Major challenges associated with this approach will pertain to the miniaturization of the device, user acceptance, and accuracy in real-world conditions.
Jaw-Mounted Strain Sensor

In 2012, Edward Sazonov and Juan Fontana proposed a method for monitoring chewing using a piezoelectric strain gauge placed on the lower jaw, right below the ear. The strain gauge is similar to that of the WearSens device (discussed next), but the placement is different; the sensor is used to detect chewing rather than swallows. As the user chews, vibrations cause voltage fluctuations at the terminals of the strain gauge, which are then processed and amplified using a custom circuit. SVMs are subsequently used to identify meal events, successfully distinguishing them from rest events and talking using a linear kernel function. However, as in the case of other devices that use inertial sensors to detect chewing and swallowing, the comfort and social acceptance aspects of this approach requires more thorough investigation.

WearSens

Figure 6 shows the WearSens necklace, which consists of a pendant-style necklace, with the pendant resting in the lower part of the neck. While an individual eats various foods, chews and swallows produce vibrations in the skin of the lower neck. Different foods can produce different vibrations, based on their unique properties. A piezoelectric sensor, mounted on the pendant, is attached such that it is in contact with the skin of the neck while the subject eats. A Bluetooth LE-enabled microcontroller samples this sensor at a rate of 20 Hz. A mobile aggregator extracts various mathematical and statistical features from these signals and applies classifiers to distinguish between a variety of foods.

The system can distinguish between water, a sandwich, and chips with an F-measure of over 90 percent while running for several days using a simple 240 mAh, 3.3V CR2032 cell battery. The system includes an Android application that aggregates sensor data and informs users of deficiencies in their diet based on quantity consumed, timing, eating pace, and skipped meals.

Review of Motion Detection

The piezoelectric sensor is passive and does not need to be powered. This sensor, along with inertial sensing mechanisms, can be sampled at rates of 20 Hz or less, which is a fraction of the rates required by audio methods.

Comfort, however, is an issue with these devices, because their placement must ensure that the sensor is in contact with the skin at all times. In addition, further work is needed to validate the device’s accuracy, social acceptance, and comfort in real-world conditions.

Online User Survey

Even the most reliable and accurate devices will fail to produce meaningful health outcomes if user adherence is low. Thus, effective wearables must be sleek, comfortable, and attractive without drawing unwanted attention. To analyze subjective factors such as comfort, convenience, and real-world feasibility, we conducted an online survey to evaluate user interest in various devices based on a brief description of major techniques.

We presented nine devices to the survey subjects, representing the most promising techniques across all the categories described in the survey. Each device was accompanied by an image and a brief description of how it’s worn and used. The system outcomes were also clearly defined in each case (that is, what the device was specifically reporting to the user). After the description and photo of each device, we presented subjects with the following questions:

- Describe your overall impression of the device, where 1 represents “I would never wear this device,” and 5 represents “I would be eager to use this device.”
- Describe your willingness to use this device in public, where 1 represents “I would never wear this or use this around others,” and 5 represents “I would have no problem wearing/using this around others.”
- Describe the usefulness of the device outputs, where 1 represents “not useful at all” and 5 represents “very useful.”

Subjects also provided their age range, gender, and the approximate location of their homes, because these factors could influence their preferences. It’s worth mentioning that the survey results are somewhat biased toward the smartwatch, which is a commercial product that has been refined over the last several years. By comparison, most of the other presented works were research prototypes. It’s likely that user impressions of these devices were heavily influenced by the appearance and product-design aspects of the images provided.

Demographics

A total of 96 subjects participated in the online survey. Of the respondents, 49 percent identified as female, 46.9 percent identified as male, and the remainder didn’t disclose their gender. In addition, 13.7 percent of subjects were under the age of 18, with the majority (51.26 percent) between the ages of 18 and 25; only 4.8 percent of participants reported their age as over 35. Overall, 73 percent of respondents claimed to be residents of the US or Canada, with

Figure 6. The WearSens necklace can detect swallows by analyzing motion of the neck as subjects eat. This necklace represents a technique for diet monitoring based on the detection of skin motion.
the next predominant demographic being Europe (18.8 percent).

**Detailed Results**

Figure 7 gives detailed survey results. We compare nine devices on the basis of respondents’ overall impressions, their comfort levels wearing or using the device in public, and their assessment of the system outputs’ usefulness. We rated these attributes on a scale of 1 to 5. The smartwatch received the highest rating overall—3.02 out of 5. The smart table and smart utensil were the next highest-scoring devices. On the other end of the spectrum, the audio headset received the lowest overall score at 1.72.

Respondents were most comfortable wearing the smartwatch by a considerable margin, receiving a score of 3.53 out of 5; the smart utensil received the next highest rating of 2.63 out of 5.

The last evaluation criterion was system outputs, which the survey described alongside photos of each device. The motivation behind this rating is that some devices claim to count calories, others aim to identify foods, and others simply recognize eating behavior. In this category, the smartwatch received the third-highest score (2.31 out of 5), due to the platform’s described limitations. The highest-rated devices were the wearable camera, smart utensil, and audio headset, with ratings of 2.90, 2.64, and 2.40, respectively.

**Analysis**

We can draw a few general conclusions from the survey results. First, several respondents expressed privacy concerns with respect to the mounted camera. One respondent stated, “Privacy and being self-conscious could be an issue,” while another said, “I have zero desire to wear a camera due to overall greater personal privacy concerns.” With respect to the audio earpiece, one respondent stated, “Could easily be hacked to use as recording devices for surveillance purposes.”

The results suggest that custom devices are generally less preferred to established hardware solutions, because the smartwatch was the overwhelmingly favorite choice in most categories. Although custom devices have a novelty factor that might appeal to some, most respondents seemed to prefer inconspicuous approaches. The smartwatch platform has both inertial sensors and a microphone, and might therefore be a viable platform for recognition of eating behavior, because a sensor fusion model can be applied to
gestures and swallow sounds for improved accuracy. Furthermore, as we mentioned, the survey tends to bias toward the smartwatch because it’s a commercially available product rather than a research prototype.

Wearable devices must provide utility, comfort, accuracy, and style to gain traction in the commercial market. Although the works presented here have demonstrated technical efficacy, user acceptance and adherence remain the biggest barriers to widespread adoption of these devices. The development of novel wearable approaches to diet monitoring will continue to be an interdisciplinary challenge, necessitating the collaboration of computer scientists, health professions, user interface designers, and HCI specialists.

ACKNOWLEDGMENTS

This work is funded by the US National Science Foundation AIR Option 1: Award Number 1312310.

REFERENCES


