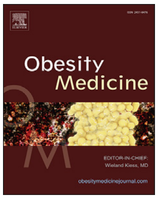
Contents lists available at ScienceDirect

Obesity Medicine

journal homepage: http://www.journals.elsevier.com/obesity-medicine



Original research

A comparison of piezoelectric-based inertial sensing and audio-based detection of swallows

Haik Kalantarian [a, *], Bobak Mortazavi [c], Nabil Alshurafa [b], Costas Sideris [a], Tuan Le [a], Majid Sarrafzadeh [a]

[a] *Department of Computer Science, University of California, Los Angeles, United States*
[b] *Department of Preventative Medicine, Northwestern University, United States*
[c] *School of Medicine, Yale University, United States*

ARTICLE INFO

Article history:
Received 21 December 2015
Accepted 12 January 2016

Keywords:
Piezoelectric sensor
Wireless health
Nutrition
Necklace

ABSTRACT

Background: Prior research has shown a correlation between poor dietary habits and countless negative health outcomes such as heart disease, diabetes, and certain cancers. Automatic monitoring of food intake in an unobtrusive, wearable form-factor can encourage healthy dietary choices by enabling individuals to regulate their eating habits.
Methods: This paper presents an objective comparison of two of the most promising methods for digital dietary intake monitoring: piezoelectric swallow sensing by means of a smart necklace which monitors vibrations in the neck, and audio-based detection using a throat microphone.
Results: Data was collected from twenty subjects with ages ranging from 22 to 40 as they consumed a variety of foods using both devices. In Experiment I, we distinguished sandwich, chips, and water. In Experiment II, we distinguished nuts, chocolate, and a meat patty. F-Measures for the audio based approach were 91.3% and 88.5% for the first and second experiments, respectively. In the piezo-based approach, F-measures were 75.3% and 79.4%.
Conclusion: The accuracy of the audio-based approach was significantly higher for classifying between different foods. However, this accuracy comes at the expense of computational overhead increased power dissipation due to the higher sample rates required to process audio signals compared to inertial sensor data.

1. Introduction

Healthy eating can reduce the risk of heart disease, stroke, diabetes, and several cancers. In 2008, medical costs associated with obesity were estimated at $147 billion, and the Centers for Disease Control (CDC) believes that the best areas for treatment and prevention are monitoring behavior and environment settings (Centers for disease control, 2014). Wireless technologies and health-related wearable devices have the potential to enable healthier lifestyle choices. These devices and systems are designed to encourage behavior modifications needed to reduce the risk of obesity and obesity-related diseases (Dorman et al., 2010).

Studies have shown that the number of swallows recorded during a day strongly correlate with weight gain on the following day (Stellar and Shrager, 1985). This provides motivation for the analysis of food intake patterns based on volume. Though many wearable devices have been designed for monitoring activity (Freedson et al., 1998; 2011; Patel et al., 2012), automatically and accurately inferring eating durations and patterns in a non-intrusive manner has been for the most part an unaddressed challenge.

Prior works have attempted to characterize eating habits through various means. Though many methods have been proposed, two of the more promising techniques include inertial-systems using piezoelectric sensors, as well as audio-based detection using throat microphones. In piezoelectric-based techniques, piezoelectric sensors, which produce a voltage in response to mechanical stress, can be used to detect movement in the skin on the lower-neck associated with swallowing. This approach differs from

* Corresponding author.
E-mail addresses: kalantarian@cs.ucla.edu (H. Kalantarian), bobak.mortazavi@yale.edu (B. Mortazavi), nabil@northwestern.edu (N. Alshurafa), costas@cs.ucla.edu (C. Sideris), tuanle@cs.ucla.edu (T. Le), majid@cs.ucla.edu (M. Sarrafzadeh).

http://dx.doi.org/10.1016/j.obmed.2016.01.003
2451-8476/© 2016 Elsevier Ltd. All rights reserved.

microphones based on piezoelectric technology: our system does not detect sound waves, instead assessing motion in the skin that results from swallows and chewing. Alternatively, audio-based techniques typically place a small microphone near the jaw or neck, and record eating noises such as chewing and swallowing. These sounds can be disambiguated from other background noises using classifiers and other signal-processing techniques. These approaches differ significantly from a perspective of comfort, practicality, convenience, power usage, and detection accuracy.

The primary novelties of our work are the description of a system in which a piezoelectric sensor is placed in the lower part of the neck for detecting swallow motions, and a comparison of this technique with audio-based monitoring using datasets derived from the same experiments. This provides a much more objective comparison of these two technologies than otherwise possible by comparing results from separate papers using different datasets and methodologies. Furthermore, we provide an evaluation of the power overhead of these techniques as a function of sample rate, computational overhead, and Bluetooth connection interval.

This paper is organized as follows. Section 2 presents related work in dietary monitoring technologies. In Section 3, we describe the hardware architecture of the two schemes, followed by algorithms in Section 4. In Section 5, we describe the experimental procedure. In Section 6, we describe experimental results. In Section 7, we describe our methods for monitoring the power and energy overhead of these techniques, which is followed by a presentation of results in Section 8. Finally, limitations and future work are described in Section 9, followed by concluding remarks in Section 10.

2. Related works

Many works have employed microphones for detecting food intake. For example, the work in Sazonov et al. (2008) uses acoustic data acquired from a small microphone placed near the bottom of the throat. Their system is coupled with a strain gauge placed near the ear. Other works suggest the use of throat microphones as a means of acquiring audio signals from throat and extracting swallowing sounds, for evaluation of dysphagia symptoms in seniors (Nagae and Suzuki, 2011; Tsujimura et al., 2010). Analyzing wave shape in the time domain or feature extraction and machine learning (Tsujimura et al., 2010) has resulted in an 86% swallow detection accuracy in an in-lab controlled environment. Similarly, the work featured in Nagae and Suzuki (2011) by Nagae et al. attempts to distinguish between swallowing, coughing, and vocalization using wavelet-transform analysis of audio data. However, identifying the volume or characteristic of food intake is not the focus of their work.

In Rahman et al. (2014), Rahman et al. present BodyBeat: a robust system for detecting human sounds. A similar work is presented by Yatani et al. in Yatani and Truong (2012). Our work differs from theirs for a number of reasons. First, we do not propose a custom hardware solution, instead employing a simple off-the-shelf throat microphone that connects directly to a mobile phone. Secondly, we emphasize classification between different foods, comparing the properties of celery, chocolate, nuts, water, chips, and sandwiches. Furthermore, we perform real-time experiments to measure the power overhead of frequency domain audio analysis, and Bluetooth 4.0 LE transmission of audio signals. Lastly, we directly compare this approach to the inertial-sensing approach on the basis of classification accuracy, and computational overhead.

In the work by Amft et al. in Amft et al. (2009), authors analyze bite weight and classify food acoustically from an earpad-mounted sensor. However, sound-based chewing recognition accuracy was low, with a precision of 60%–70%. In Amft (2010), the authors present a similar earpad-based sensor design to monitor chewing sounds. Food grouping analysis revealed three significant clusters of food: wet and loud, dry and loud, soft and quiet. An overall recognition accuracy of over 86.6% was achieved. Some studies have reached accuracy rates of 91.7% in an in-lab controlled environment using neural networks with false positives of 9.5% (Aboofazeli and Moussavi, 2004). A more recent study using support vector machines have been able to reach swallow detection accuracies of up to 84.7% in an in-lab setting (Sazonov et al., 2010). These devices are mounted very high in the upper trachea, near the laryngopharynx. In Passler and Fischer (2011), Pler, et al. proposed a system geared towards patients living in ambient assisted living conditions and used miniature electret microphones which were integrated into a hearing aid case, and placed in the ear canal. Our prior work described in Kalantarian et al. (2014a) also provided a foundation for spectrogram-based analysis of audio signals. A similar approach for analyzing bioacoustic signals using spectrograms was also presented by Pourhomayoun et al. in Pourhomayoun et al.

A “smart tablecloth” was presented in 2015 by Bo Zhou et al. in Zhou et al. (2015). The system detects eating behavior on solid surfaces (such as tables), based on changes in the pressure distribution of these tables during the eating process. The tablecloth was a matrix of pressure sensors based on a carbon polymer sheet, which changes its electrical resistance in response to electrical force. At the corners of the tablecloth, force-sensitive resistors (FSRs) are installed with the primary purpose of determining weight, rather than spatial density. Features extracted from the FSRs, as well as the pressure-sensitive tablecloth, are analyzed using classifiers such as decision trees to distinguish between various eating-related actions such as stirring, scooping, and cutting. Based on the ratio of different actions performed, the authors were able to distinguish between four different meal types with high accuracy. Furthermore, changes in the average pressure values from the data stream were associated with a decrease in the remaining amount of food on the table, which was used to estimate food weight with an error of approximately 16.62%.

The E-Button was presented in 2014 by Professor Mingui Sun at the University of Pittsburgh (Sun et al., 2014). In this work, Sun et al. propose a chest-mounted button with an embedded camera that among other applications, can be applied to the domain of dietary monitoring. 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 acquired data is transmitted to a smartphone using Bluetooth or WiFi. The E-Button operates by taking photos at a preset rate, thereby recording the entire eating process. Using image processing techniques, the utensils (such as a plate or bowl) are detected. Subsequently, the food items are identified based on color, texture, and other heuristics. Using this information and additional DSP techniques, volume figures can be calculated for each food, which is converted to a Calorie count using a public domain database that equates volume and food type to Calories. Evaluation of 100 foods was conducted, and the error was approximately 30% for 85% of the foods, which were regularly shaped. However, irregularly shaped food was not detected with high accuracy.

Several prior works have attempted to detect swallow disorders using piezoelectric sensors. The work by Toyosato et al. in Toyosato et al. (2007) used a Piezoelectric Pulse Transducer to detect food bolus passage through the esophagus. In Ertekin et al. (1996), Ertekin et al. used piezoelectric sensors to evaluate dysphagia symptoms in a study with thirty normal subjects and 66 dysphagia patients. The authors concluded that piezoelectric sensors can be applied successfully towards objective evaluation of oropharyngeal dysphagia. Another example is our prior work in nutrition monitoring in Kalantarian et al. (2014b), in which we propose monitoring eating habits by placing a piezoelectric sensor in the lower trachea. The work by Alshurafa et al. in Alshurafa et al. (2014) provides a foundation for our piezoelectric classification algorithm using spectrograms. In Miyaoka et al. (2011), Miyaoka et al. used piezoelectric sensor signals were able to detect the volume of tea swallowed based on the waveforms acquired from the sensor, after GLM-ANOVA analysis.

3. Hardware architecture

3.1. Audio-based approach

In order to analyze the volume and consistency of a meal, audio samples are acquired while an individual is eating, using a commercial throat microphone placed freely in the lower part of the neck. A primary advantage of this method is comfort, as unlike the piezoelectric sensor, it is not necessary to have skin contact at all times using this approach. The microphone was resting loosely around the throat near the lower collarbone. The particular microphone used in this study is the Hypario Flexible Throat Mic Microphone Covert Acoustic Tube Earpiece Headset, which is connected directly to the mobile phone's audio input port using a 3.5 mm male audio cable, and picks up audio signals from swallows, through the air. Commercially available audio-recording technology was used to acquire the audio recordings from the microphone.

3.2. Piezo-based approach

A piezoelectric sensor, sometimes known as a vibration sensor, produces a voltage when subjected to physical strain. By placing a piezoelectric sensor against the throat, the motion of the skin during a swallow is represented in the output of the sensor, when sampled at frequencies as low as 5 Hz. The NIMON necklace, presented by Kalantarian et al. in Kalantarian et al. (2014b), describes a non-invasive, wearable device capable of detecting swallows by placement of a vibration sensor near the lower trachea. During a swallow event, muscular contractions cause skin motion, which pushes the vibration sensor away from the body and towards the fabric of the necklace, generating a unique output voltage pattern. The skin motion during a swallow is quite small and requires very high sensitivity to detect. Therefore, the long and thin design of a piezoelectric strip very well suited for this application, in which the necklace clamps the sensor to the skin and causes it to bend slightly during a swallow (Fig. 1).

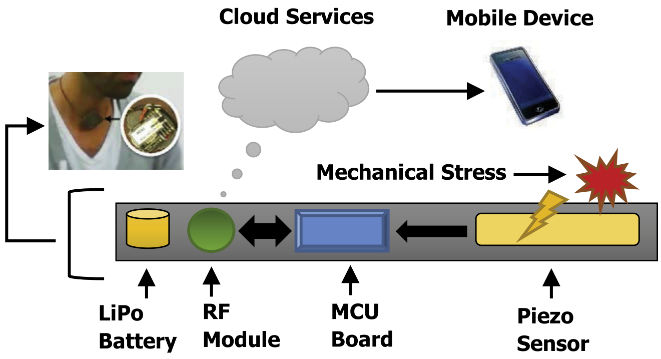


Fig. 1. Systems architecture of the piezoelectric sensor-based necklace used for swallow detection.

Raw data is sampled from the vibration sensor at a rate of 20 Hz, and a windowing algorithm computes the standard deviation of the values in each window (typically sized 20). Subsequently, the peaks are identified using on a thresholding technique, which typically correspond with swallows. The various steps in data processing are shown in Fig. 2, with the raw value visible at the top. The noticeable

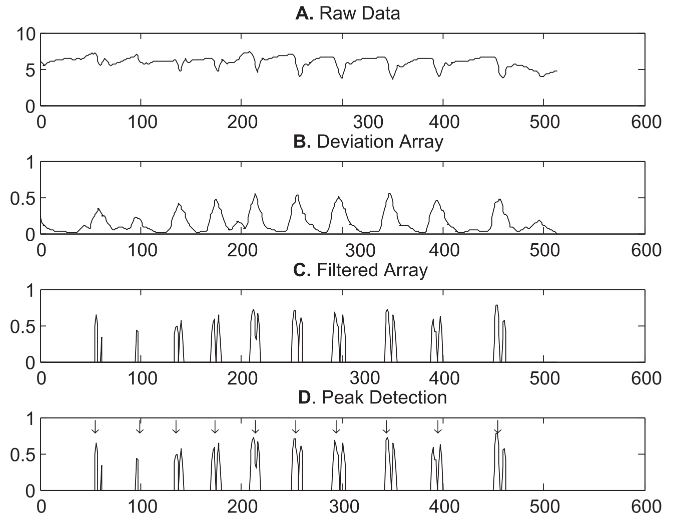


Fig. 2. This graph shows the signal processing flow used by the Nimon necklace to identify swallows.

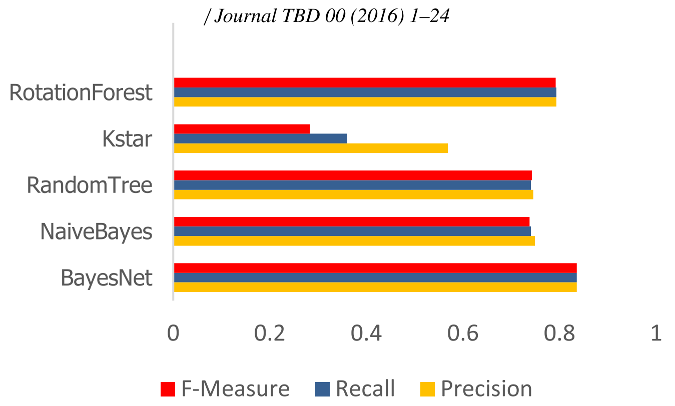


Fig. 3. This figure provides a comparison of the accuracies of various classifiers for the throat microphone, based on their precision, recall, and f-measure.

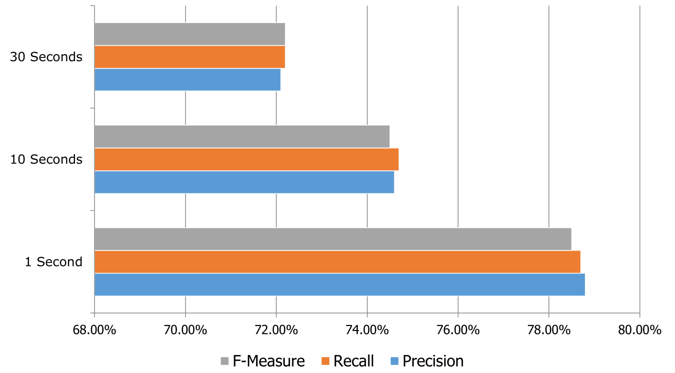


Fig. 4. This graph shows how varying window sizes affected the classification accuracy of audio-based signals. The 1-s window size appeared to have the highest classification accuracy.

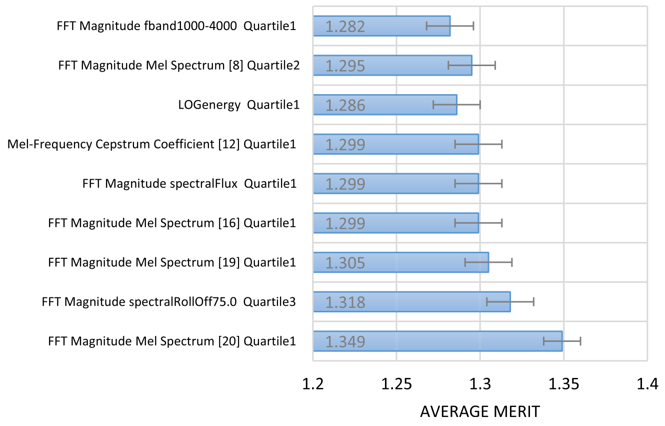


Fig. 5. This graph shows the major features for audio-based classification using the openSMILE toolkit, for audio-based classification. The InfoGain Attribute Evaluation tool was used for feature selection (opensmile faq).

dips in the waveform generally correspond with swallows. After initial data acquisition, the data is smoothed using a moving-average low-pass filter with a span of 5, to reduce the impact of noise. Subsequently, a sliding window of length 9, corresponding with .45 s of data, is applied with a maximum overlap (shifted one point at a time). These numbers were experimentally determined to be optimal for preserving the critical features of the waveform based on simulations. The original implementation of the NIMON necklace used Bluetooth LE to transmit all raw data to an Android phone for processing; the algorithm does not run on the embedded hardware, which is powered by a small lithium-polymer battery (Figs. 3–5).

The piezoelectric sensors can be used for more advanced classification activities, beyond counting swallows. Because the piezoelectric sensor is capable of detecting motions beyond swallows, the detection of consistent chewing between swallows is a reliable indicator that a solid food is being consumed, while several swallows, with no chewing between them, may indicate that a liquid is being consumed. Generally speaking, the foods with different textures produce different patterns of vibrations in the neck, as a result of the varying amounts of jaw strength necessary for chewing, as well as varying speeds at which foods are eaten, chewed, and swallowed. Analysis of the statistical features associated with the raw data sampled from the piezoelectric sensor can reveal the food being consumed, within a limited subset.

4. Algorithms

4.1. Audio feature extraction and classification

The Munich open Speech and Music Interpretation by Large Space Extraction toolkit, known as openSMILE (Eyben et al., 2010), is a feature extraction tool intended for producing large audio feature sets. This tool is capable of various audio signal processing operations such as applying window functions, FFT, FIR filterbanks, autocorrelation, and cepstrum. In addition to these techniques, openSMILE is capable of extracting various speech related features and statistical features. A partial list of extracted features is shown in Tables 1 and 2, respectively. Other audio-based features include frame energy, intensity, auditory spectra, zero crossing rate, and voice quality. After data is collected from a variety of subjects eating several foods, feature selection tools can be used to identify strong features that are accurate predictors of swallows and bites for various foods, while reducing the dimensionality by eliminating redundant or weakly correlated features (Table 3).

Table 1
Partial List of openSMILE Speech Features from (opensmile faq).

Speech-related features		
Signal energy	Loudness	Mel/Bark/Octave Spectra
MFCC	PLP-CC	Pitch
Voice quality	Formants	LPC
Line spectral pairs	Spectral shape	CENS and CHROMA

Table 2
Partial list of openSMILE statistical features from (opensmile faq).

Speech-related features		
Means	Extremes	Moments
Segments	Samples	Peaks
Zero crossings	Quadratic regression	Percentiles
Duration	Onset	DCT Coefficient

A custom script concatenates all recorded audio clips from throat microphone experiments into one large audio file, and splits them into smaller clips of length n, based on the input parameters. This allows the extraction of eating clips without a-priori knowledge of when the swallows take place, for experimental variety. The feature extraction tool then extracts a large feature set of over 6500 attributes per clip, which provides the classifier with enough information to distinguish between the different categories. The InformationGain Attribute Evaluation tool then reduces the feature sets by selecting those which have minimum redundancy and maximum correlation to the defined classifier outcomes.

4.2. Piezoelectric feature extraction and classification

A spectrogram is a visual representation of the frequency spectrum over time, and is an ideal representation for extracting distinguishing features in many classification problems. A spectrogram is typically generated using a short-time Fourier transform (STFT) with a fixed window size, the squared magnitude of which yields the spectrogram. Fundamentally, a spectrogram allows easy identification of changes in the frequency spectrum of a signal, over time. This is significant because eating certain foods produce vibrations in different frequency ranges, based on the texture of the food and the amount of chewing involved. A spectrogram can provide a relatively straightforward representation of changes in the frequency distribution over time as subjects eat, which can reveal the distinguishing characteristics of various foods.

The spectrogram is calculated from the time signal $x(t)$, as shown in Eq. (1) using the short-time Fourier transform (STFT).

$$STFT\{x(t)\} \equiv X(n,\omega) = \sum_{t=-\infty}^{\infty} x[t]\omega[t-n]e^{-j\omega t}. \qquad (1)$$

$x(t)$ is multiplied by a window function for a short period of time. The data is divided into frames F_i, which overlap. Each frame is Fourier transformed, and the result is added to a matrix that records the magnitude and phase of each point in time and frequency. A Hamming window was used of varying lengths of $w = 32$, 64, and 128, with an FFT length of $nfft = 32$, 64, and 128, and an of overlap of 25%, 50%, 75%, and no overlap. We set the dynamics range to 50 dB. Each spectrogram is defined by a matrix $P \in R^{m \times k}$, where m is the number of bins in the time domain, and k is the number of bins in the frequency domain. P represents the power spectral density.

Once a spectrogram is generated for each swallow, we found an optimal division of the spectrogram images into 14 bins along the frequency domain and another 16 bins along the time domain, for a

Table 3
This table shows a list of the most important features extracted from the audio spectrogram as well as their accompanying descriptions.

Extracted feature	Description
$\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} a_{xy}}{m*n}$	The average value of amplitude within a sample window.
$\sqrt{\frac{\sum_{x=1}^{m}\sum_{y=1}^{n}(a_{xy}-\mu)}{m*n}}$	The standard deviation of amplitude within a sample window
$\frac{\sum_{x=1}^{n} a_{zx}}{n}$ $for\{z\|1 \le z \le m\}$	Average of the various frequency bins. Each frequency range is extracted separately as an independent feature.
$\sqrt{\frac{\sum_{x=1}^{n}(a_{zx}-\mu)}{n}}$ $for\{\|1 \le z \le m\}$	The standard deviation of a frequency range over a period of time, for every frequency bin.

Table 4
Feature table for piezoelectric sensor feature extraction.

Mean	Geometric mean	Std. Dev.
Skewness	Mean of standardized Z-scores	IQR
Kurtosis	Harmonic mean	Rank corr.
Range	Median absolute deviation	Partial corr.

total of 30 bins. We then calculate statistical features on each bin, to generate a feature vector V_i for each swallow. Table 4 lists the main features that were calculated for each bin, which generates a total of $s = 360$ features per spectrogram swallow. The motivation for the use of these features is based on the work by Alshurafa et al. in Alshurafa et al., which demonstrated the superiority of the spectrogram-based statistical approach over alternatives such as matching pursuit, and scalogram-based Gabor wavelets.

5. Experimental procedure

5.1. Piezoelectric sensor data collection

Two experiments were performed to validate the efficacy of our algorithm in accurately detecting swallows and recognizing eating patterns using statistical features collected from a spectrogram. To prevent bias in the classification results between each class label in the training set, we randomly select an equal number of swallows across categories. We also perform leave-one-out cross validation and report the results.

In the first experiment data was collected on ten subjects, two female and eight male with ages ranging between 20 and 40 years of age. We placed the necklace around their neck so that the sensor was loosely touching the skin. The necklace tightness was adjusted such that each subject was comfortable wearing the device. We placed the necklace centered between their right and left clavicle right above the sternum and asked the subject to eat the following foods, one at a time: a chicken salad or tuna salad sandwich, a small handful of Pringles potato chips, and a cup of 9oz water. No specific instructions were given about the manner in which the food was consumed, though subjects were under observation during the data collection process which may have increased the pace of eating beyond what would otherwise be typical.

In the second experiment we increased the number of subjects to twenty, eight female and twelve male, ages 20–40 years. The subjects each consumed a meat-like veggie patty, a handful of mixed nuts, and two small Snickers chocolate bars. We ensured that the portion sizes were identical from one subject to another. The subjects were asked to push a button every time they swallowed; this helped us further annotate the data in order to provide truth labels for the dataset.

5.2. Audio data collection

Our data collection includes data from 20 individuals using a throat microphone placed near the bottom of the neck. The moments at which food was swallowed were indicated by pressing a push button which added an annotation to the associated log file. In the original data collection, ten subjects were instructed to eat two identical sandwiches (3-inch and 6-inch), and drink two cups of water (9 fl. oz and 18 fl. oz), and eat a small handful (approximately 3) of Pringles chips. They were also instructed not to eat, swallow, or speak for a brief period in order to acquire signals corresponding to silence, or background noise. Subsequently, 189 audio samples were extracted from the recordings. The next phase of experimentation employed twenty subjects, who were given a small portion of nuts, chocolate, a vegetarian meat-substitute patty. The foods were consumed sequentially, in that order. Over 50 additional samples were manually extracted from this experiment, though some data was corrupted. These recordings formed the basis of the algorithm design and experimental evaluation. For evaluation of classification results, leave-one-out cross validation was used.

The data collection took place in a lab environment; people can be faintly heard speaking in the background, and the microphone occasionally recorded doors closing and nearby footsteps. In most audio classification works, ambient noises can interfere with the signal and decrease classification accuracy. This issue is partially rectified by placing the throat microphone in the lower part of the neck. This microphone placement emphasizes swallow sounds, as they are in much closer proximity to the device than ambient noises. Furthermore, most commercial throat microphones contain active circuitry for filtering out these ambient signals. These factors make throat microphones particularly well-suited for the task of recognizing eating behavior from chew and swallow sounds. Our approach for detecting ingestion, despite the presence of background noises, is similar to the evaluation conducted by Kalantarian et al. in Kalantarian and Sarrafzadeh. The experiments presented in this paper suggested that spectrogram-based feature extraction techniques are relatively resilient against the ambient noises, as the frequency bands associated with external sounds are typically not selected features by the classifier models and therefore do not significantly affect the classification results.

6. Evaluation

6.1. Audio-based classification results

Table 5 shows the accuracy of acoustic swallow detection for three food types: water, sandwich, and chips using 1-s samples. Chips in particular had the highest classification accuracy, with a recall and precision of 100%, based on 50 samples. The precision and recall of water was lower, at 87.7% and 86%, respectively. This is

Table 5
Audio: Confusion matrix (Random Forest) using a 1-s window.

Swallow type	Predicted outcome			Recall
	Water	Sandwich	Chips	
Water	43	7	0	86%
Sandwich	6	44	0	88%
Chips	0	0	50	100%
Precision	87.7%	86.3%	100%	

partially because of the automated method of dividing the water sample clips. Because drinking water has no chewing, and very little information between swallows, it can be difficult to classify an audio clip that happens to be taken between swallows. Table 6 shows the classification accuracy of nuts, chocolate, and the veggie-patty. Of note are the poor results for nuts, with a recall of 78.0%.

Because classification was among four foods, rather than binary classification, results are quite promising. However, it is clear that the algorithm may not scale well if tested on a larger number of food types (ie. 50–100). Therefore, it is desirable for future systems to create very broad categories based on highly generalizable features. For example, the work by Amft et al. in Amft (2010) describes a scheme which generalizes food into three significant clusters of food: “wet and loud,” “dry and loud,” and “soft and quiet.” Further research is necessary to determine typical nutritional qualities associated with these types of foods.

6.2. Piezoelectric-based classification results

According to our classification results, using spectrogram-based features on a signal from a piezoelectric sensor can distinguish between liquid and solid swallows with high accuracy using the Random Forest Classifier (with n = 100 trees), which yielded the optimal results for all three experiments. Best results were achieved using a window size of 32, an FFT length of 32, and an overlap of 50%. Generally, it was observed that the Random Forest Classifier consistently outperforms other well-known classifiers, even when distinguishing between solids, achieving an F-measure of 75.29% in the first experiment (water, sandwich, chips) and 79.44% in the second (nuts, chocolate, patty).

The Random Forest classifier (Liaw and Wiener, 2002) is an ensemble-learning decision tree classifier that, unlike most other tree-based classifiers which split each node based on the best subset of features, instead uses the best subset of predictors randomly chosen at that node. This unique property of the Random Forest classifier has been shown to make it relatively robust to overfitting, and has compared favorably against other classifiers such as support vector machines and neural-network classifiers. A more detailed investigation of the properties of this classifier can be found in the work by Leo Breiman in Breiman (2001).

The Bayesian Network and kNN classifier resulted in a 72.6% and 65.4% F-measure, respectively. Table 8 provides the confusion matrix for the Random Forest Classifier for the first experiment, while Table 7 shows results for the second. Though results using the piezoelectric sensor were generally strong, especially in light of the very low sample rate of 20 Hz (compared to 44,000 Hz for the microphone), audio-based classification has higher accuracy. However, approach this comes at the expense of computational and power overhead.

Table 6
Audio: Confusion matrix (Random Forest) using a 1-s window.

Swallow type	Predicted outcome			Recall
	Nuts	Chocolate	Patty	
Nuts	39	6	5	78.0%
Chocolate	0	49	1	98.0%
Patty	4	1	45	90.0%
Precision	90.6%	87.5%	88.2%	

Table 7
Piezoelectric sensor: Confusion matrix (Random Forest).

Swallow type	Predicted outcome			Recall
	Water	Sandwich	Chips	
Water	43	3	4	86.0%
Sandwich	2	37	11	74.0%
Chips	5	12	33	66.0%
Precision	86.0%	71.1%	68.7%	

Table 8
Piezoelectric sensor: Confusion matrix (Random Forest).

	Predicted outcome			
Swallow type	Nuts	Chocolate	Patty	Recall
Nuts	35	11	4	70.0%
Chocolate	5	40	5	80.0%
Patty	2	4	44	88.0%
Precision	83.3%	72.7%	83.0%	

7. Energy modeling methods

To realize the intended goal of minimizing burden, it is desirable for wearable devices to remain powered for weeks, or months, without interruption. Required nightly charging or frequent coin-cell battery replacement can be considered an addition burden to the user, which is likely to lower long-term compliance rates. Therefore, modern wearable devices are carefully designed to minimize power usage, and run for days or even months. For example, the Misfit Shine activity monitor claims a battery life of four months (Misfit wearables faq). Other wearables devices such as the Jawbone UP24 claim their devices can sustain seven days of continuous use (Jawbone up technical specifications). Another important drawback of wearable devices with high power requirements are their reliance on large batteries, which may impact the comfort, convenience, and aesthetic appearance of the device. Therefore, it is necessary for wearable devices to carefully factor energy efficiency in their design. This is particularly true in the case of nutrition monitoring devices which are typically mounted in the throat or jaw area.

We now discuss a comparison of power consumption for both schemes. Table 9 presents the hardware components used for power evaluation of the devices. Note that the original microcontroller board, the Arduino-based RFDuino, has been substituted for an MSP430 variant for the superior simulation environment and embedded low-power architecture. The MSP430 has 5 different power modes: Active Mode, and Low Power mode (LP) 1–4. Active mode is used when executing algorithms, while the low power

Table 9
A list of system components used in evaluation.

Hardware components (NIMON necklace)	Description
Nordic nRF8002	Bluetooth 4.0 Module
LIS3DH Accelerometer	Accelerometer
MSP430G2744	Microcontroller
CR2032 Coin-Cell	Battery for powering device

modes supply a clock signal to various peripherals. LP3 disables the CPU, MCLK, and SMCLK. Note that MCLK is the main system clock of the MSP430, used by the CPU. SMCLK is the sub-main clock used by other peripherals such as timers. To conserve energy, the device is configured to alternate between Active Mode and LP3. This can be seen in Fig. 6; the device briefly enters Active Mode to acquire, process, and transmit data, returning to LPM3 immediately thereafter to conserve energy and maximize battery life (Figs. 7–10).

The selected Bluetooth transceiver is the Nordic nRF8002, commonly used in several commercial wearables including the Misfit Shine. Power evaluation of the Bluetooth Transceiver was provided by Nordic Semiconductor's nRFGo Studio software (version 1.17), which simulates power demands and battery lifetime based on many different parameters such as data size, encryption, connection intervals, advertising intervals, and transmit strength. Real-time power debugging of the MSP430 microcontroller was provided by the EnergyTrace++ technology in Texas Instruments Code Composer Studio (CCS) Version 6.0, which provides real-time information about processor state and power consumption. The data in the following sections corresponds with the piezoelectric-based designs unless otherwise specified. However, most of the results can be generalized to draw conclusions between these two schemes. For example, both techniques require sampling, buffering, and transmission. Specific to the audio-based approach is the FFT algorithm, for which we provide an analysis.

8. Power evaluation

Eq. 2 shows a representation of the energy needed to acquire a sample from the piezoelectric sensor via the on-chip ADC, process the data, and transmit via Bluetooth 4.0 to mobile phone. Eq. 3 shows the power consumption of the system (P) in terms of the sample rate and the energy required to acquire, process, and transmit a sample, which are the major sources of power loss in both the audio-based and piezoelectric-based systems.

$$E_n = E_{adc} + E_{proc} + E_{tx} \tag{2}$$

$$P = \sum_{a=1}^{f} \left[E_{adc} + E_{proc} + E_{tx}\right] = \sum_{a=1}^{f} E_n \tag{3}$$

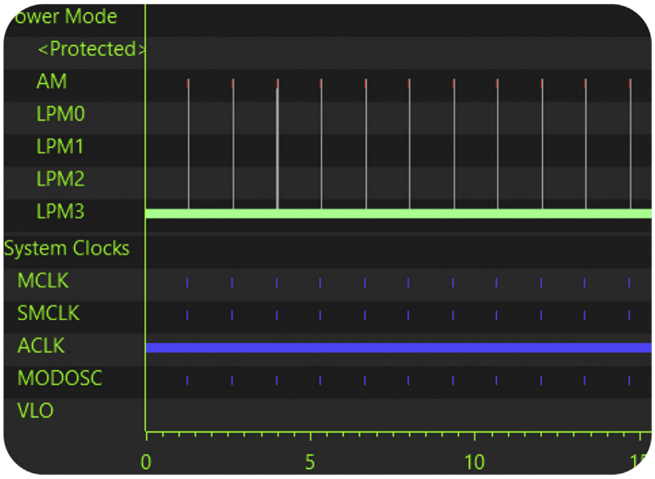


Fig. 6. This figure shows the state transitions of the MSP430 microcontroller, which occur on an interrupt callback. in LPM3, all clocks besides ACLK (necessary for the timer) are disabled.

Window	Sample Rate				
Size	**1 Hz**	**4 Hz**	**8 Hz**	**16 Hz**	**32 Hz**
4	.05 mW	.06 mW	.06 mW	.07 mW	.07 mW
10	.05 mW	.06 mW	.06 mW	.07 mW	.08 mW
20	.05 mW	.06 mW	.06 mW	.08 mW	.10 mW
100	.06 mW	.08 mW	.10 mW	.16 mW	.26 mW

Fig. 7. Relationship between window size, sample rate, and mean current drawn on MSP430 mcu with on-board swallow detection.

Window	Sample Rate				
Size	**1 Hz**	**4 Hz**	**8 Hz**	**16 Hz**	**32 Hz**
4	.05 mW	.05 mW	.05 mW	.06 mW	.06 mW
10	.05 mW	.05 mW	.05 mW	.06 mW	.06 mW
20	.05 mW	.05 mW	.05 mW	.06 mW	.06 mW
100	.05 mW	.05 mW	.05 mW	.06 mW	.06 mW

Fig. 8. Relationship between window size, sample rate, and mean current drawn on MSP430 mcu when swallow detection is offloaded to the mobile phone.

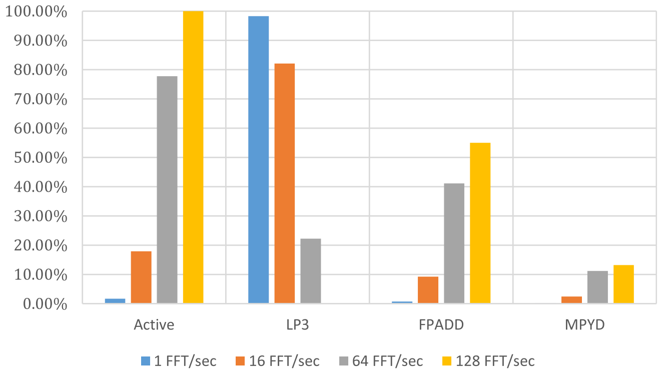


Fig. 9. This figure shows the percentage of time the MSP430 spent in various modes, as a result of 16-point FFT operations performed at different rates.

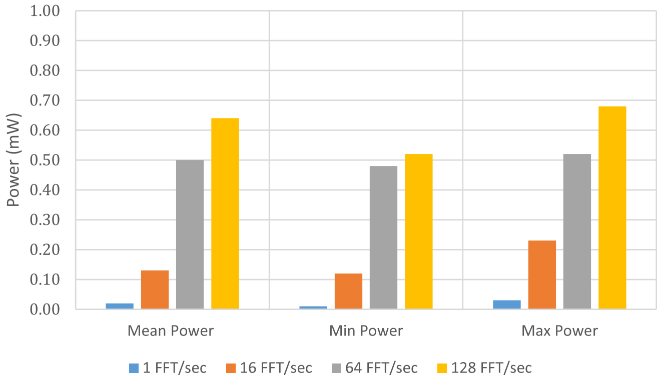


Fig. 10. This figure shows the power dissipation of the microcontroller when 16-point FFT operations are performed at various rates.

8.1. *Sample rate, window size, and power*

Table 7 shows the relationship between sample rate and power, at several different window sizes. Note that the window size refers to how many samples a given value acquired from the piezoelectric sensor is compared against, to calculate the variance of the data. Recall that the sample rate is associated with the frequency at which samples are acquired from the piezoelectric sensor, buffered locally, and processed. Therefore, a high sample rate not only incurs local processing overhead based on the implementation of the swallow detection algorithm, but also directly relates to the usage of the on-chip Analog-Digital converter of the MSP430. Table 8 shows similar numbers for the case when data processing is offloaded to the mobile phone. In this case, the device simply acquires raw data and transmits it immediately, and the contributions to the reported power are AD conversion and transmission. This table shows that increasing the sample rate between 1 Hz and 32 Hz has a quite limited effect on power usage, since the AD converter is in fact designed for significantly higher sample rates. In comparison with Table 7, we conclude that the high sample rate consumes more power largely by increasing the number of data points that must be processed.

Table 7 also shows the relationship between window size and power, at several different sample rates. It should be noted that the window size is associated with the piezoelectric-based algorithm only, in our evaluation. It refers to the size of the sliding average window used in the peak detection algorithm. A key observation is that increasing the window size does not have a substantial effect on power usage, for low sample rates (Kalantarian et al., 2015). The NIMON necklace has been tested at sample rates of 10 Hz and 20 Hz, with window sizes between 4 and 10 samples. The window size appears to be a significant contributor to low device lifetime only at very high sample rates. Also, note that in Table 8, offloading the processing from the necklace to the smart phone decouples the window size parameter from the current usage of the device.

8.2. *Power vs. connection interval*

Power results for various connection intervals are shown in Table 10. These numbers assume a fixed bandwidth based on data acquisition at 20 Hz. Therefore, higher connection intervals also have higher payloads. Results show that average power ranges from 1.2 mW at a connection interval of 50 ms, to 0.026 mW at a connection interval of 3 s. Compared to the extra .01 mW for performing swallow detection locally, it is clear this is more optimal to run detection algorithms locally for anything responsive enough to be considered real time user feedback (10 s or less). Clearly, the connection interval is a significant contributor to overall power dissipation, and this should be increased whenever possible.

8.3. *Energy evaluation of fourier-based algorithms*

In this section, we briefly cover the device lifetime challenges using acoustic sensors. In Amft et al. (2009), Amft et al. were able to recognize chewing with high precision using a sample rate of 44 kHz, which is quite high for an embedded application. This is especially the case when compared to the approach of the piezoelectric necklace in which sample rates as low as 10 Hz have been validated. Though clearly, a 44 kHz audio signal will have more information for classification than a much smaller sample rate, the implications of battery life will be substantial. In another work by Amft et al. in Amft et al. (2005), audio data was also acquired at 44 kHz and processed with a 512-point FFT. However, prior works (Brochetti et al., 1992; Hi et al., 1988), have shown spectral energy from potato chips to be primarily between 0 and 10 kHz, with highest amplitude frequency ranges between 1 and 2 kHz. Based on Nyquist-Shannon sampling theory, this conservative estimate would require a sample rate of 4 kHz. Though this is well within the specifications of the MSP430 ADC unit, it far exceeds the sample rate of 10–20 Hz, as required by the vibration sensor. The overhead of acquiring and transmitting up to 400× more data, not to mention the computational overhead of the FFT in comparison to the simple windowing algorithm described in this paper, makes the vibration sensor a far better choice.

Table 10
Average transmit power for a fixed bandwidth (20 Hz) at 3.7 V.

T_{conn} (ms)	Payload (bytes)	Avg current (μA)	Avg power (μW)
50	6	328.37	1214.97
100	12	260.70	964.59
200	24	195.66	723.94
300	36	99.98	369.92
500	60	41.98	155.32
1000	120	19.90	73.62
3000	360	7.18	26.56

For evaluation, a 16-point FFT algorithm was implemented on the MSP430 microcontroller. The power debugging did not evaluate the impact of high-sample ADC units; we instead measured the power usage of the FFT algorithm itself. The sampling, buffering of results, and transmission would incur significant additional overhead. In a continuous signal processing application such as this, the FFT algorithm must be run periodically to analyze frequency domain features of incoming data. Table 10 shows results for four different FFT rates (1, 16, 64, and 128). For example, an FFT rate of 16 represents an operating mode in which a 16-point FFT is evaluated 16 times per second (therefore processing 256 samples per second). The system was designed such that the MSP430 microcontroller would immediately enter LPM3 (low power mode 3) to conserve energy between FFT operations. The table shows that average power is quite high for high FFT rates such as 128 (corresponding to 2048 samples/second, or a sample rate of 2 kHz); mean power approached .64 mW, without considering the additional overhead of sampling at 2 kHz, buffering the results, and transmitting to mobile phone. Table 9 shows that as the frequency of FFT operations increases, the MSP430 microcontroller spends a higher percentage of its time in Active Mode, in which power is not conserved. At the rate of 128 FFTs/second, the device spends all of its time in Active Mode, and 55% of its time performing a floating point add operation. This suggests that the microcontroller is unable to perform FFT operations at the requested rate, since it should enter LPM3 after completion. At this rate, the power dissipation is over ten times greater than that of a vibration-sensor based system with a sample rate of 8 Hz and a window size of 20, as shown in Table 7.

9. Limitations and future works

There are several limitations to the current work that should be addressed in the future. First, the evaluation of the power usage of an audio system is simplified, as a more robust scheme would vary the sample rate based on the ambient noise and therefore consume very little power in quiet environments. The implications of related energy-saving technologies throughout an entire day of use should be evaluated.

Secondly, user receptiveness to these schemes should be quantified based on surveys and focus groups. The continuous monitoring of audio signals may make some individuals uncomfortable due to the privacy issues associated with this technique; the recorded audio signal can contain speaking in addition to eating sounds, despite the placement of the device on the lower neck area. A potential solution would be to remove the ability of the device to wirelessly transmit the acquired signals to a mobile phone. Instead, the device could process the data locally, communicating feedback to the user using LEDs, vibrations, or a simple display. A more detailed study of privacy issues in ubiquitous systems is provided by Langheinrich et al. in Langheinrich (2001), and Hong et al. in Hong and Landay (2004).

Though the piezoelectric-sensor technique does not raise the same privacy concerns, user receptiveness to device comfort may be a significant issue. This is a result of the placement of the device, which requires that the piezoelectric sensor be placed against the skin of the neck. Though it is not necessary for the necklace to be placed very tightly, it must still be in contact during eating, which may be uncomfortable for some individuals and potentially impractical for obese subjects with significant adipose fat tissue; assessment of eating habits for obese subjects has not been evaluated in existing work.

10. Conclusion

In this paper, we provide an analysis of two emerging techniques for monitoring eating habits in consumer electronics applications: using piezoelectric sensors placed in the bottom of the neck, and throat microphones for analysis of audio signals associated with the ingestion of food. Results suggest that audio-based classification has somewhat higher accuracy, particularly for dry foods such as chips and nuts. However, we show that the audio-based approach incurs a power overhead approximately ten times greater than the vibration-sensor system, which may have broad implications with respect to device form factor, user acceptance and comfort.

Conflict of interest

There is no conflict of interest.

Acknowledgments

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

References

Aboofazeli, M., Moussavi, Z., 2004. Automated classification of swallowing and breadth sounds. Conf. Proc. IEEE Eng. Med. Biol. Soc. 5, 3816–3819.

Alshurafa, N., Kalantarian, H., Pourhomayoun, M., Sarin, S., Liu, J., Sarrafzadeh, M., 2014. Non-invasive monitoring of eating behavior using spectrogram analysis in a wearable necklace. In: IEEE EMBS Healthcare Innovations & Point of Care Technologies (HIPT).

Alshurafa, N., Kalantarian, H., Pourhomayoun, M., Sarin, S., Liu, J., Sarrafzadeh, M., Recognition of nutrition-intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor, IEEE Sens. J.

Amft, O., 2010. A wearable earpad sensor for chewing monitoring. In: Sensors, 2010 IEEE, pp. 222–227. http://dx.doi.org/10.1109/ICSENS.2010.5690449.

Amft, O., Stger, M., Lukowicz, P., Trster, G., 2005. Analysis of chewing sounds for dietary monitoring. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (Eds.), UbiComp 2005: Ubiquitous Computing, Vol. 3660 of Lecture Notes in Computer ScienceSpringer, Berlin Heidelberg, pp. 56–72. http://dx.doi.org/10.1007/115512014. URL. http://dx.doi.org/10.1007/115512014.

Amft, O., Kusserow, M., Troster, G., 2009. Bite weight prediction from acoustic recognition of chewing. IEEE Trans. Biomed. Eng. 56 (6), 1663–1672. URL. http://dblp.uni-trier.de/db/journals/tbe/tbe56.html.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brochetti, D., Penfield, M.P., Burchfield, S.B., 1992. Speech analysis techniques: a potential model for the study of mastication sounds. J. Texture Stud. 23 (2), 111–138. http://dx.doi.org/10.1111/j.1745-4603.1992.tb00515.x. URL. http://dx.doi.org/10.1111/j.1745-4603.1992.tb00515.x.

Centers for Disease Control and Prevention: Adult Obesity Facts, 2014. URL. http://www.cdc.gov/obesity/data/adult.html.

Dorman, K., Yahyanejad, M., Nahapetian, A., Suh, M.-k., Sarrafzadeh, M., McCarthy, W., Kaiser, W., 2010. Nutrition monitor: a food purchase and consumption monitoring mobile system. In: Phan, T., Montanari, R., Zerfos, P. (Eds.), Mobile Computing, Applications, and Services, vol. 35. Springer, Berlin Heidelberg, pp. 1–11. http://dx.doi.org/10.1007/978-3-642-12607-91.

Ertekin, C., Aydodu, I., Yceyar, N., 1996. Piecemeal deglutition and dysphagia limit in normal subjects and in patients with swallowing disorders. J. Neurol. Neurosurg. Psychiatry 61 (5), 491–496. http://dx.doi.org/10.1136/jnnp.61.5.491 arXiv. http://jnnp.bmj.com/content/61/5/491.full.pdf+html. http://jnnp.bmj.com/content/61/5/491.abstract. URL.

Eyben, F., Wöllmer, M., Schuller, B., 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the International Conference on Multimedia, MM'10. ACM, New York, NY, USA, pp. 1459–1462. http://dx.doi.org/10.1145/1873951.1874246. URL. http://doi.acm.org/10.1145/1873951.1874246.

Freedson, P.S., Melanson, E., Sirard, J., 1998. Calibration of the computer science and applications, Inc. Accelerom. Med. Sci. Sports Exerc. 30 (5), 777–781.

Freedson, P.S., Lyden, K., Kozey-Keadle, S., Staudenmayer, J., 2011. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. J. Appl. Physiol. 111 (6), 1804–1812.

Hi, W.E.L., Deibel, A.E., Glembin, C.T., Munday, E.G., 1988. Analysis of food crushing sounds during mastication: frequency-time studies. J. Texture Stud. 19 (1), 27–38. http://dx.doi.org/10.1111/j.1745-4603.1988.tb00922.x. URL. http://dx.doi.org/10.1111/j.1745-4603.1988.tb00922.x.

Hong, J.I., Landay, J.A., 2004. An architecture for privacy-sensitive ubiquitous computing. In: Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services. ACM, pp. 177–189.

Jawbone up technical specifications, http://jawbone.com/store/buy/up24.

H. Kalantarian, M. Sarrafzadeh, Audio-based Detection and Evaluation of Eating Behavior Using the Smartwatch Platform, Elsevier Computers in Biology and Medicine.

Kalantarian, H., Alshurafa, N., Pourhomayoun, M., Sarin, S., Le, T., Sarrafzadeh, M., 2014. Spectrogram-based audio classification of nutrition intake. In: IEEE EMBS Healthcare Innovations & Point of Care Technologies (HIPT).

Kalantarian, H., Alshurafa, N., Sarrafzadeh, M., 2014. A wearable nutrition monitoring system. In: IEEE Body Sensor Networks.

Kalantarian, H., Alshurafa, N., Pourhomayoun, M., Sarrafzadeh, M., 2015. Power optimization for wearable devices. In: IEEE Percom: WristSense.

Langheinrich, M., 2001. Privacy by designprinciples of privacy-aware ubiquitous systems. In: Ubicomp 2001: Ubiquitous Computing. Springer, pp. 273–291.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R. News 2 (3), 18–22.

Misfit wearables faq, http://www.misfitwearables.com/supportl.

Miyaoka, Y., Ashida, I., ya Kawakami, S., Tamaki, Y., Miyaoka, S., 2011. Generalization of the bolus volume effect on piezoelectric sensor signals during pharyngeal swallowing in normal subjects. J. Oral Biosci. 53 (1), 65–71. http://dx.doi.org/10.1016/S1349-0079(11)80037-X. http://www.sciencedirect.com/science/article/pii/S134900791180037X. URL.

Nagae, M., Suzuki, K., 2011. A Neck Mounted Interface for Sensing the Swallowing Activity Based on Swallowing Sound. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 5224–5227. http://dx.doi.org/10.1109/IEMBS.2011.6091292.

opensmile faq, http://www.audeering.com/research/opensmile.

Passler, S., Fischer, W., 2011. Acoustical method for objective food intake monitoring using a wearable sensor system. In: Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 266–269, 5th International Conference on, 2011.

Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M., 2012. A review of wearable sensors and systems with application in rehabilitation. J. NeuroEng. Rehabil. 9 (1), 21.

Pourhomayoun, M., Dugan, P., Popescu, M., Clark, C., Bioacoustic signal classification based on continuous region processing, grid masking and artificial neural network, arXiv preprint arXiv:1305.3635.

Rahman, T., Adams, A.T., Zhang, M., Cherry, E., Zhou, B., Peng, H., Choudhury, T., 2014. Bodybeat: a mobile system for sensing non-speech body sounds. In: Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'14. ACM, New York, NY, USA, pp. 2–13. http://dx.doi.org/10.1145/2594368.2594386. URL. http://doi.acm.org/10.1145/2594368.2594386.

Sazonov, E., Schuckers, S., Lopez-Meyer, P., Makeyev, O., Sazonova, N., Melanson, E.L., Neuman, M., 2008. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. Physiol. Meas. 29 (5), 525. URL. http://stacks.iop.org/0967-3334/29/i=5/a=001.

Sazonov, E.S., Makeyev, O., Schuckers, S., Lopez-Meyer, P., Melanson, E.L., Neuman, M.R., 2010. Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. IEEE Trans. Biomed. Eng. 57 (3), 626–633.

Stellar, E., Shrager, E.E., 1985. Chews and swallows and the microstructure of eating. Am. J. Clin. Nutr. 42 (5 Suppl. l), 973–982.

Sun, M., Burke, L.E., Mao, Z.-H., Chen, Y., Chen, H.-C., Bai, Y., Li, Y., Li, C., Jia, W., 2014. ebutton: a wearable computer for health monitoring and personal assistance. In: Proceedings of the 51st Annual Design Automation Conference. ACM, pp. 1–6.

Toyosato, A., Nomura, S., Igarashi, A., Ii, N., Nomura, A., 2007. A relation between the piezoelectric pulse transducer waveforms and food bolus passage during pharyngeal phase of swallow. Prosthodont. Res. Pract. 6 (4), 272–275. http://dx.doi.org/10.2186/prp.6.272.

Tsujimura, H., Okazaki, H., Yamashita, M., Doi, H., Matsumura, M., 2010. Non-restrictive measurement of swallowing frequency using a throat microphone. IEEJ Trans. Electron. Inf. Syst. 130, 376–382. http://dx.doi.org/10.1541/ieejeiss.130.376.

Yatani, K., Truong, K.N., 2012. Bodyscope: a wearable acoustic sensor for activity recognition. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, ACM, pp. 341–350.

Zhou, B., Cheng, J., Sundholm, M., Reiss, A., Huang, W., Amft, O., Lukowicz, P., 2015. Smart table surface: a novel approach to pervasive dining monitoring. In: Pervasive Computing and Communications (PerCom), pp. 155–162. http://dx.doi.org/10.1109/PERCOM.2015.7146522. IEEE International Conference on, 2015.