2014 Health Innovations and Point-of-Care Technologies Conference
Seattle, Washington USA, October 8-10, 2014

Non-Invasive Monitoring of Eating Behavior using Spectrogram Analysis in a Wearable Necklace

Nabil Alshurafa, Haik Kalantarian, Mohammad Pourhomayoun, Shruti Sarin, Jason J. Liu, Majid Sarrafzadeh

***Abstract*— Food intake levels, hydration, chewing and swallowing rate, and dietary choices are all factors known to impact one's health. This paper presents a novel wearable system in the form of a necklace, which aggregates data from an embedded piezoelectric sensor capable of detecting skin motion in the lower trachea during ingestion. We propose an algorithm based on spectrogram analysis of piezoelectric sensor signals to accurately distinguish between food types such as liquid and solid, hot and cold drinks and hard and soft foods. The necklace transmits data to a smartphone, which performs the processing of the signals, classifies the food type, and provides visual feedback to the user to assist the user in monitoring their eating habits over time. Experimental results demonstrate high classification accuracy of the proposed method, and validate the use of a spectrogram in extracting key features representative of the unique swallow patterns of various foods.**

I. Background and Related Works

Healthy eating is associated with reduced risk for many diseases, including several of the leading causes of death: heart disease, some cancers, stroke, and diabetes [1]. However much of the wireless technology developed and used in the market is focused primarily on exercise and physical activity [2], [3], [4]. In this paper we deploy a novel system that attempts to classify food types from a device in the form of a necklace.

Automatically and accurately inferring eating durations and patterns in a non-intrusive manner has been for the most part an unaddressed challenge. Current technologies for eating pattern detection are either inaccurate or exhibit low rates of adherence to using the technology, due to one or more of these shortcomings: 1) they infer eating indirectly from, for example, hand movements or food images [5], [6]; 2) they are non-pervasive requiring manual data entry or user involvement in capturing data [7]; 3) they are non-wearable, bulky, invasive, or semi-invasive [8]; 4) they exhibit low accuracy in detecting swallows and distinguishing food types [9]. In this paper, we focus on classifying food types using a novel piezoelectric-based design of a necklace [10] that is not tightly worn around the neck, but rather hangs loosely like a true necklace right above the sternum.

In this paper, we deploy a classification technique using a spectrogram-based approach which considers time and frequency-based components in tandem. A spectrogram, often used for speech recognition, and other countless applications is a visual representation of the frequency spectrum over time generated using a short-time Fourier transform (STFT) with a fixed window size, the squared magnitude of which yields the spectrogram. Spectrograms have been applied to countless research problems pertaining to the analysis of acoustic signals. Examples include the identification of whale sounds, speech recognition, and pattern recognition in genome sequences [11], [12]. However, their utility in analyzing piezoelectric sensor data has not been adequately explored. The primary novelty of our work is the application of spectrograms for analysis of piezoelectric sensor data in the realm of detection and classification of food ingestion.

II. Nutrition monitoring necklace design

Our nutrition monitoring system comprises two main components: piezoelectric-based sensor technology; and a smartphone application that performs data processing, user guidance and feedback. The smartphone application performs swallow detection, feature extraction and classification to detect swallows. This section describes the sensor technology and user guidance and feedback. Section III will further discuss the classification algorithms implemented on the smartphone.

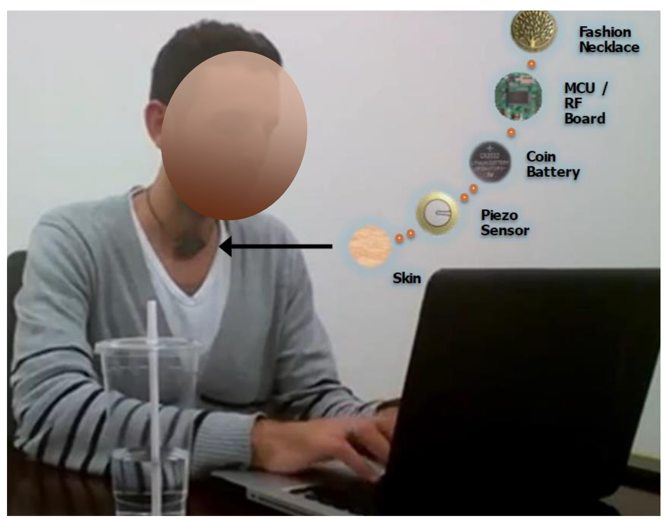

Fig. 1. This figure shows a subject wearing the necklace while working. The necklace comprises a piezoelectric sensor, coin battery, MCU/RF board and a fashionable cover, while maintaining contact between the sensor and skin.

A piezoelectric sensor, also known as a vibration sensor, produces a voltage when subjected to physical strain. By placing a piezoelectric sensor against the throat, the muscle contraction and motion of the skin during a swallow is represented in the output voltage of the sensor, when sampled at frequencies as low as 5 Hz. Our necklace features a

Department of Computer Science, University of California, Los Angeles
E-mail: {nabil,kalantarian,mpourhoma,sarin,jasonliu,majid}@cs.ucla.edu

thin, lightweight piezoelectric vibration sensor attached to the inside of the necklace, along with a small microcontroller board capable of sampling the sensor and transmitting the data to a mobile phone via Bluetooth. The hardware is powered by a lightweight lithium-polymer battery. Figure 1 depicts each component of the wearable necklace.

The microcontroller board samples the voltage of the vibration sensor at a rate of 20Hz, converting the voltage to a digital signal using the on-chip A/D converter. The data is then buffered and transmitted to a mobile phone. This Arduino-compatible board features a Bluetooth 4.0 LE transceiver, based on the RFD22301 SMT module. The embedded processor is an ARM Cortex M0 with 256kB of flash memory and 16kB of RAM.

This system includes a mobile phone application for data reporting and visualization. The application reports on estimates of portion size, dehydration levels, rate of consumption, and eating habits like skipping meals and binge eating.

III. Algorithm

A. Swallow Detection

Figure 2 illustrates a simplified version of the algorithm implemented on the mobile phone, which is used to detect swallows based on data acquired from the vibration sensor and received via Bluetooth. The data is buffered locally until a sufficient number of samples have been acquired. Subsequently, a sliding window is applied to generate a new waveform representing the standard deviation of the original data. The swallows are represented in the resulting waveform as peaks, while they may correspond to either peaks or troughs in the original data.

The algorithm then proceeds to smooth the waveform by applying a Savitzky-Golay convolution filter, which yields clearly visible peaks representative of each original swallow, while removing noise from the signal. Consequently, the number of swallows is identified by counting the number of peaks, provided there is sufficient spacing between swallows.

B. Spectrogram

Once a swallow is detected, a spectrogram is generated centered around each swallow. The spectrogram is calculated from the time signal $x(t)$ using the short-time Fourier transform (STFT). $x(t)$ is multiplied by a window function for a short period of time. The data is divided into frames, 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. The spectrogram is the resulting 3-dimensional plot of the energy of the frequency content of a signal as it changes over time [13]. For our window function, we used a Hamming window of length $w = 32$, with an FFT length of $nfft = 32$, and an of overlap of 50%.

Figure 3 provides an illustration showing a sample swallow spectrogram for three food types water, chips and sandwich. The distinguishing attributes of these foods are visible. For example, chips and sandwich swallows contain

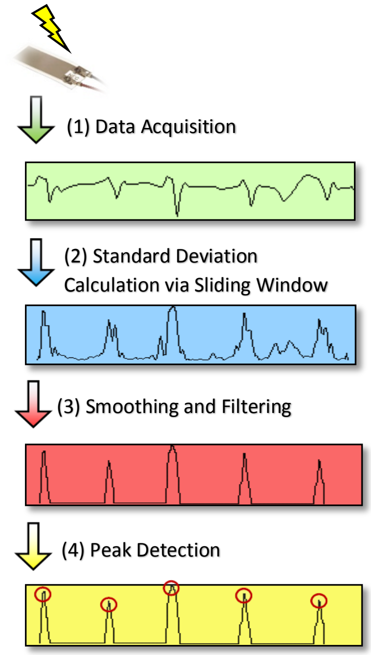

Fig. 2. This figure shows the swallow detection process, whereby a sliding window is applied to the signal to generate a waveform representing the standard deviation of the signal, after smoothing, peaks are detected and identified as potential swallow regions.

more high frequency components than water swallows due to the effect of chewing. Distinguishing between chips and sandwich swallows, though significantly more challenging, is captured by our statistical feature extraction methodology.

C. Feature Selection and Classification

Once we generate a spectrogram for each swallow, we divide the spectrogram image into 14 bins along the frequency domain and another 16 bins along the time domain, for a total of 30 bins. We then calculate statistical features on each bin, to generate a feature vector for each swallow. We also generate features on the entire spectrogram image. Table I lists the main features that were calculated for each bin, which generates a total of 372 features per swallow.

TABLE I
Feature Table

Mean	Geometric Mean	Std. Dev.
Skewness	Mean of Standardized Z-Scores	IQR
Kurtosis	Harmonic Mean	Rank Corr.
Range	Median Absolute Deviation	Partial Corr.

The conventional feature selection algorithms usually focus on specific metrics to quantify the relevance and/or redundancy in the feature set in order to find the smallest subset of features that provides the maximum amount of useful information for prediction. Thus, the main goal of feature selection algorithms is to eliminate redundant or irrelevant features in a given feature set. Applying an effective feature selection algorithm not only decreases the computational

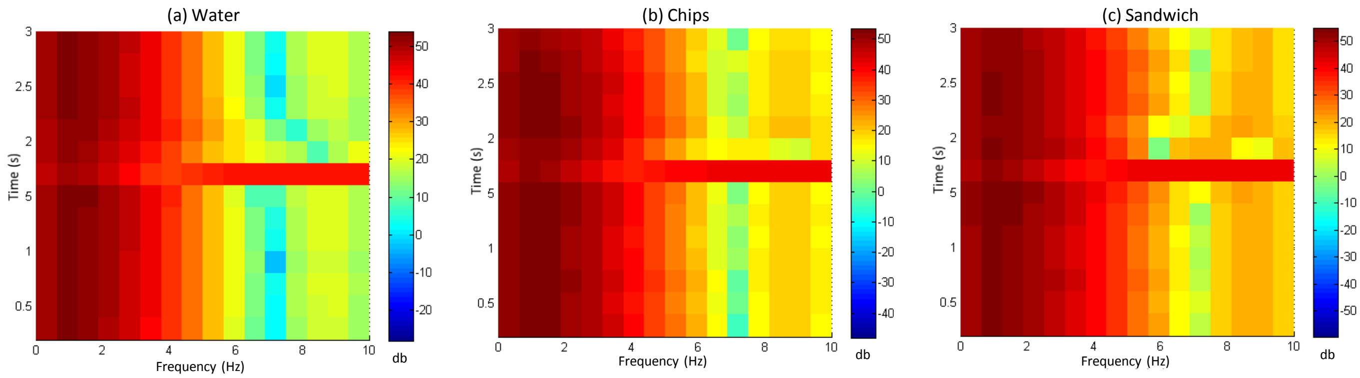

Fig. 3. Figure a shows the spectrogram for a water swallow over a 3 second window. Figure b and Figure c show the spectrogram for chips and sandwich, respectively. The distinction between water and solid is evident in the spectrogram which exhibits lower frequency components, while the sandwich exhibits more higher frequency components. Finer differences are captured when we extract statistical features on each spectrogram.

complexity of the system by reducing the dimensionality and eliminating the redundancy, but also increases the performance of the classifier by deleting irrelevant and confusing information.

The two well-known feature selection categories are the filter and wrapper methods. Filter methods use a specific metric to score each individual feature (or a subset of features together), and are usually fast and much less computationally intensive. Wrapper methods usually utilize a classifier to evaluate feature subsets in an iterative manner according to their predictive power [14]. We applied the wrapper method, testing on multiple combinations of feature subsets and classifiers including: kNN, Bayesian Network, Random Forest. The optimal feature subset and classifier combination is selected to run in real time. Figure 4 provides an illustration of the system architecture, where an optimal feature subset and classifier is trained to distinguish between food types.

IV. Experimental Setup

An experiment was performed on a total of 20 subjects to validate the efficacy of our algorithm in accurately detecting and classifying swallows. 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.

Each subject consumed three types of solid food: a meat-like veggie patty, a handful of mixed nuts, and two fun-sized snicker bars. They also consumed two types of liquid, an 8 ounce glass of room temperature water and 8 ounces of hot tea. We compare the classifiers ability to distinguish between liquid and solid as well as different textures, temperatures, and consistencies. 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. To prevent bias in the classification results between each class label, we randomly selected an equal number of swallows across categories. We also performed

TABLE II
Confusion matrix for Liquids vs. Solids

	Predicted Outcome		
Swallow Type	Liquid	Solid	Recall
Liquid	207	33	86.3%
Solid	31	209	87.1%
Precision	87.0%	86.4%	

TABLE III
Confusion matrix for Hot Tea vs.Water

	Predicted Outcome		
Swallow Type	Hot Tea	Water	Recall
Hot Tea	92	8	92.0%
Water	12	88	88.0%
Precision	88.5%	91.7%	

Leave One Subject Out Cross Validation. We test each classifier's ability to distinguish between different food types and select the optimal classifier.

V. Results And Discussion

The feature selection algorithm that consistently performed best in combination with the classifiers was the correlation-based feature subset selection algorithm [15]. The classifier that provided the best results was also the Random Forest Classifier with number of trees set to 100. Results show the ability of the algorithm to accurately distinguish between liquid and solids as shown in Table II.

The Liquid class label precision and recall was 87.0% and 86.3%, respectively. The Solid class label produced a precision and recall of 86.4% and 87.1%, respectively, which shows that the collected spectrogram features are good discriminants between liquids and solids. The Random Forest Classifier resulted in a 86.6% F-measure compared to 77% and 78.1% F-measures of Bayesian Networks and kNN (with k=3 yielding the best result), respectively.

When testing the ability of the algorithm to distinguish between hot tea and water, our results show that the Random Forest Classifier resulted in a 90% F-measure, compared to 69.0% and 84.0% F-measures of Bayesian Networks

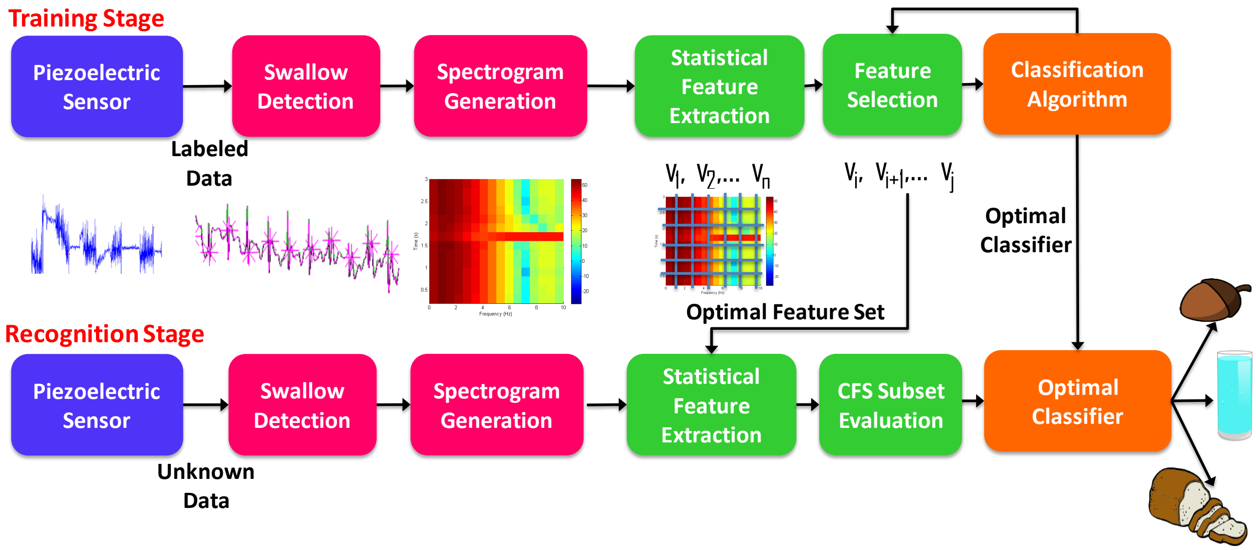

Fig. 4. System Architecture.

TABLE IV
CONFUSION MATRIX FOR DISTINGUISHING SOLIDS

	Predicted Outcome			
Swallow Type	Nuts	Chocolate	Patty	Recall
Nuts	36	10	4	72.0%
Chocolate	5	41	4	82.0%
Patty	2	5	43	86.0%
Precision	83.7%	73.2%	84.3%	

and kNN (with k=3 yielding the best results), respectively. The Hot Tea precision and recall was 88.5% and 92.0% respectively. The Water precision and recall was 91.7% and 88.0%, respectively. Table III provides the confusion matrix between the Hot Tea and Water class label. The Random Forest Classifier resulted in a 90.0% F-measure

Our findings show that it is challenging to distinguish between solids. As seen in Table IV the Random Forest Classifier consistently outperforms other well-known classifiers, even when distinguishing between solids, achieving an F-measure of 80%. The Bayesian Network and kNN classifier resulted in a 72.6% and 65.4% F-measure, respectively. Table IV provides the confusion matrix for the Random Forest Classifier. While there exists an infinite number of food concoctions people often maintain a regular regimen, and such a system can become customized to distinguish between different foods within a subject's diet.

VI. CONCLUSION

In this paper we performed classification of swallows using statistical features collected from spectrograms generated from piezoelectric sensor signals. Our results show promise in using spectrogram analysis in combination with piezoelectric sensors as opposed to audio sensors. We have developed and tested a necklace prototype which has shown the ability to successfully distinguish between liquids and solids using Random Forest Classifier with 100 trees resulting in an F-measure of 87%. We show methods of distinguish between hot and cold drinks with an F-measure of 90%. We also show potential for distinguishing between solid food types. As a future work, we intend to expand classification to more food types, while testing in a more free-living environment.

REFERENCES

[1] "Dietary guidelines for americans," United States Department of Agriculture, Tech. Rep. 32877, 2010.

[2] N. Alshurafa, W. Xu, J. J. Liu, M.-C. Huang, B. Mortazavi, M. Sarrafzadeh, and C. K. Roberts, "Robust human intensity-varying activity recognition using stochastic approximation in wearable sensors," in *BSN*, 2013, pp. 1–6.

[3] N. Alshurafa, W. Xu, J. Liu, M. Huang, B. Mortazavi, C. Roberts, and M. Sarrafzadeh, "Designing a robust activity recognition framework for health and exergaming using wearable sensors," vol. PP, no. 99, 2013, pp. 1–1.

[4] N. Alshurafa, J.-A. Eastwood, M. Pourhomayoun, S. Nyamathi, L. Bao, B. Mortazavi, and M. Sarrafzadeh, "Anti-cheating: Detecting self-inflicted and impersonator cheaters for remote health monitoring systems with wearable sensors," in *BSN*, 2014, pp. 92–97.

[5] Y. Dong, A. Hoover, and E. Muth, "A device for detecting and counting bites of food taken by a person during eating," in *Bioinformatics and Biomedicine, 2009. BIBM '09. IEEE International Conference on*, Nov 2009, pp. 265–268.

[6] N. Yao, R. Sclabassi, Q. Liu, and M. Sun, "A video-based algorithm for food intake estimation in the study of obesity," in *Bioengineering Conference, 2007. NEBC '07. IEEE 33rd Annual Northeast*, March 2007, pp. 298–299.

[7] M. Jain, G. R. Howe, and T. Rohan, "Dietary assessment in epidemiology: Comparison of a food frequency and a diet history questionnaire with a 7-day food record," *American Journal of Epidemiology*, vol. 143, no. 9, pp. 953–960, 1996.

[8] F. G. D'Ottaviano, T. A. Linhares Filho, H. M. T. d. Andrade, P. C. L. Alves, and M. S. G. Rocha, "Fiberoptic endoscopy evaluation of swallowing in patients with amyotrophic lateral sclerosis," *Brazilian Journal of Otorhinolaryngology*, vol. 79, pp. 349 – 353, 06 2013.

[9] O. Amft, "Methods for detection and classification of normal swallowing from muscle activation and sound," in *Proceedings of the First International Conference on Pervasive Computing Technologies for Healthcare, ICST*, ser. Proceedings of the First International Conference on Pervasive C, 0 2006.

[10] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "A wearable nutrition monitoring system," in *BSN*, 2014, pp. 75–80.

[11] D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, 2000.

[12] D. Sussillo, A. Kundaje, and D. Anastassiou, "Spectrogram analysis of genomes," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 29–42, Jan. 2004.

[13] J. Flanagan, *Speech analysis synthesis and perception*, ser. Kommunikation und Kybernetik in Einzeldarstellungen. Springer-Verlag, 1972.

[14] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[15] M. A. Hall, "Correlation-based feature subset selection for machine learning," Ph.D. dissertation, University of Waikato, Hamilton, New Zealand, 1998.