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

Spectrogram-Based Audio Classification of Nutrition Intake

Haik Kalantarian, Nabil Alshurafa, Mohammad Pourhomayoun, Shruti Sarin, Tuan Le, Majid Sarrafzadeh
University of California, Los Angeles
Computer Science Department

***Abstract*—Acoustic monitoring of food intake in an unobtrusive, wearable form-factor can encourage healthy dietary choices by enabling individuals to monitor their eating patterns, maintain regularity in their meal times, and ensure adequate hydration levels. In this paper, we describe a system capable of monitoring food intake by means of a throat microphone, classifying the data based on the food being consumed among several categories through spectrogram analysis, and providing user feedback in the form of mobile application. We are able to classify sandwich swallows, sandwich chewing, water swallows, and none, with an F-Measure of 0.836.**

Keywords—nutrition; swallow detection; spectrogram

I. INTRODUCTION

Obesity is a pervasive crisis around the world, especially in the United States, and its impact is ever-increasing. More than 2/3rd of American adults are overweight or obese, and the health problems associated with obesity cost the United States an estimated $148 billion in 2008 [8][10]. Being overweight or obese has countless well-documented adverse affects on an individual's well being, such as increased risk of coronary heart disease, high blood pressure, stroke, diabetes, and various forms of cancer [9].

The last decade has seen the rise of sensor-based wearable health-monitoring systems geared towards improving health [1][5][17]. Studies have revealed that such hardware systems and software applications in the realm of dietary monitoring can encourage users to meet their nutrition goals. Subjects who participated in such programs were shown to lose substantial amounts of body fat through the feedback and user guidance made possible by these platforms [13]. However, increasing physical activity levels is just one component necessary for weight loss, and maintaining healthy eating habits remains an important factor. For example, studies have revealed that various trends in eating frequency, the number of skipped meals, and the timing of food consumption can be associated with the prevalance of obesity [14]. This necessitates the design of practical, lightweight, wearable, wireless sensor-based systems capable of monitoring nutrition intake. Furthermore, merely identifying the number of swallow events may not be sufficient for all use cases, as many recommended dietary techniques suggest individuals to increase their liquid consumption while reducing their intake of other foods. This motivates the design of the system and algorithms described in this paper.

Here, we present an acoustic technique for detection and classification of swallow events in a mobile, wearable platform. The system is packaged in the form of a throat microphone, and an associated Android application which rapidly samples audio data associated with chewing and swallowing various foods. Through spectrogram analysis, feature extraction, and the application of several learning algorithms, basic classification can be performed, which provides more insight into the nutrition intake of the subject. The system architectural flow is shown in Figure 1, from audio acquisition to user guidance. Figure 2 shows how wearable sensors, mobile applications, and web services can be integrated into a complete dietary intake monitoring system.

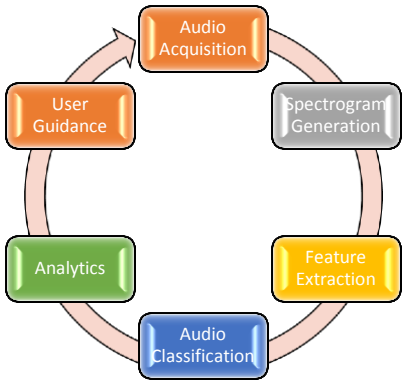

Figure 1 - This figure illustrates the system flow

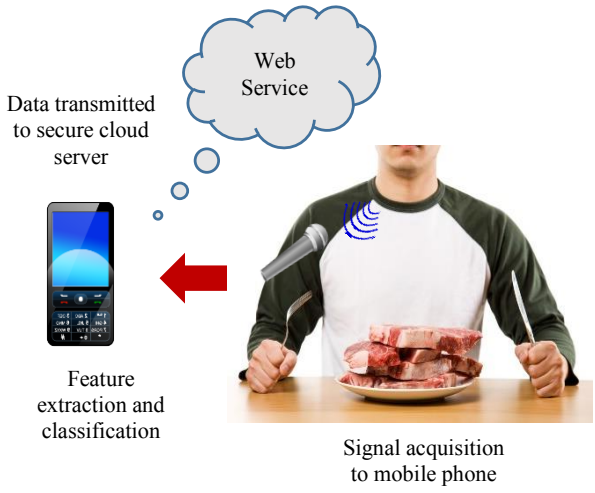

Figure 2 –A microphone placed near the throat passively monitors eating habits in real time, relaying the acquired information to a mobile phone for processing. The data is subsequently uploaded to a secure cloud server.

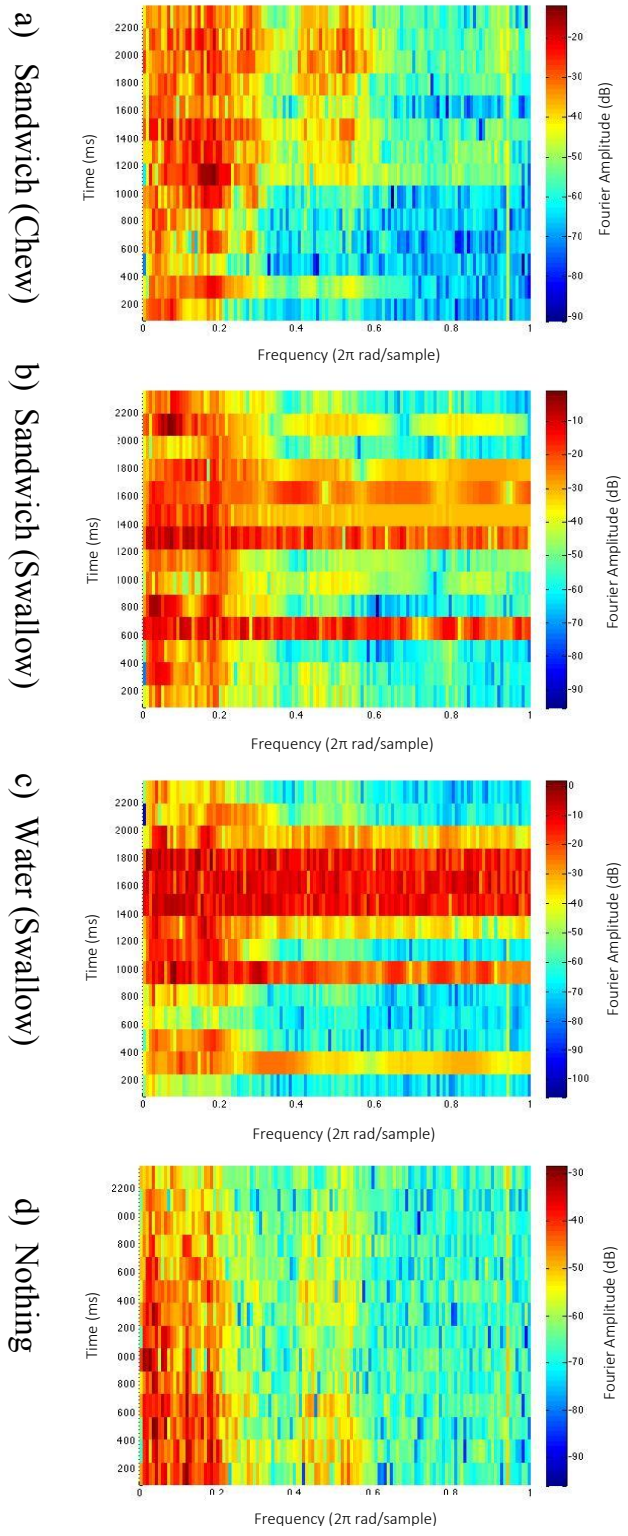

Figure 3 - The spectrograms for various food types reveal their distinguishing attributes, which can be used for classification.

II. Related Work

The authors of [4] and [5] attempt to assess dysphagia, which is difficulty swallowing frequently experienced by seniors, using acoustic swallow detection. Their work is promising for dysphagia, but they do not target weight loss and nutrition intake. However, their results are encouraging because they reveal that a throat microphone can perform swallow-detection without significant influence caused by outside noise.

In [1], the authors propose a method for detecting food intake using a piezoelectric strain gauge sensor placed in the lower jaw, and are able to classify periods of sitting, speaking, and eating with an accuracy of over 80%. While the results are promising, we believe the conspicuous sensor placement may make regular use a challenge, outside of a clinical environment.

The work presented in [17] features a technique for acoustic detection of swallow events. Frequency domain analysis is performed on different time windows, and relevant features are extracted. Using an SVM classifier, the detection rate of individual swallow events was above 80%. This work motivates our approach, suggesting that acoustic swallow detection can be extended to classification between different food types, which is essential to ensure the practicality of any passive nutrition-monitoring system.

Another interesting work on acoustic swallow detection is presented in [6], with basic classification between swallows and breath sounds performed using a feedforward neural classifier. They perform a manual inspection of their classification results using a spectrogram, which is a basis for the feature extraction technique for food classification used in our work. Another work which applies spectrograms for audio classification using machine learning algorithms is presented in [7], though their analysis is limited to detecting swallows, rather than performing classification.

III. Experimental Setup

Prior to algorithm development, data was collected from ten subjects while eating, 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. Each subject was instructed to eat two identical sandwiches (3-inch and 6-inch), and drink two cups of water (9 oz and 18 oz). Furthermore, each subject was instructed not to eat, swallow, or speak, for a brief period. Subequently, 189 audio samples were extracted from the recordings which corresponded to sandwich swallows, sandwich chews, water swallows, and silence. These recordings formed the basis of the algorithm design and experimental evaluation.

IV. Algorithms

A. Spectrogram Generation

An algorithm was developed to classify the four types of collected audio recordings. This was achieved by generating a spectrogram corresponding with each audio clip. A spectrogram is a visual representation of the frequency spectrum over time, and the spectrograms of most sounds have several distinguishing features. 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.

For spectrogram generation, a Hamming window was applied of length 1024, and an FFT length of 4ms (64 samples) based on extracted half-second audio samples centered on the swallow. No overlap was used between neighboring segments. Figure 3 shows spectrograms for audio clips corresponding with sandwich swallows, sandwich chews, water swallows, and no action. The distinguishing attributes of these audio recordings are clearly visible. For example, water swallows contain more high frequency components than sandwich swallows and are shorter in duration. Sandwich chewing features primarily low-frequency components, while the state of neither chewing nor swallowing reveals relatively unchanging frequency distributions over time, with few high frequency components.

B. Feature Extraction

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{2n} \\ \vdots & a_{32} & a_{33} & a_{3n} \\ a_{m1} & a_{m2} & a_{m3} & a_{mn} \end{bmatrix}$$

Figure 4 – The spectrogram of each window are represented by a matrix with dimensions m x n; each column represents the frequency range and each row represents the duration of the FFT window.

Figure 4 shows a matrix representation of a spectrogram, which is used to extract distinguishing features for each audio clip, to be used for classification. By extracting features across all frequency bands, the classification tool is able to determine which frequencies are most characteristic of different types of swallows. Based on this notation, Table 1 shows a list of some of the most important features.

C. Classification

Several machine learning algorithms were applied to the extracted feature set to classify between the four categories of: chewing (sandwich), swallow (sandwich), swallow (water), and nothing. These algorithms include Rotation Forest, Random Forest, Bayesian Network Classifier, and K-Star.

TABLE I – EXTRACTED FEATURES

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.
$\frac{\sum_{x=1}^{n} a_{zx}}{n}$ $for\{z \mid 1 \le z \le m\}$	Average of the various frequency bins. Each frequency range is extracted separately as an independent feature.
$\sqrt{\frac{\sum_{x=1}^{n}(a_{zx}-\mu)}{n}}$ $for\{z \mid 1 \le z \le m\}$	The standard deviation of a frequency range over a period of time, for every frequency bin.

Table 1 – This table shows a list of the most important features extracted from the spectrogram as well as their accompanying descriptions.

TABLE II – EXPERIMENTAL RESULTS

		Predicted Class			
		Chew	**Nothing**	**Sandwich Swallow**	**Water Swallow**
Actual Class	**Chew**	39	3	2	0
	Nothing	6	44	0	0
	Sandwich Swallow	4	0	34	7
	Water Swallow	0	5	4	41

V. Results

A. Classification Accuracy

Table II shows the accuracy of swallow detection for all four scenarios. The average recall and precision are both 0.836, using the Bayesian Network Classifier. The resulting F-Measure is therefore 0.836.

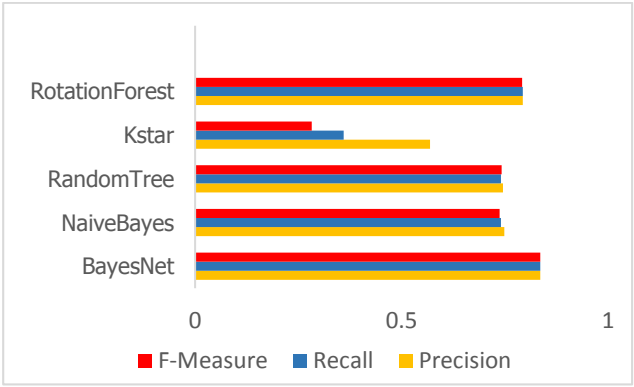

Figure 5 – This figure provides a comparison of the accuracies of various classifiers, based on their precision, recall, and f-measure.

VI. Conclusion

Spectrograms, while typically used for speech processing, speech training, the study of phonetics, as well as countless other applications, appears to be very well suited for classifying different food types by exploiting the unique features found in audio recordings of swallows of various food types. In this paper, we are able to classify between sandwich swallows, water swallows, sandwich chews, and none, using a Bayesian classifier and features extracted from an audio spectrogram. The resulting F-Measure for classification is 0.836, based on 189 collected samples.

References

[1] Sazonov, E. S., & Fontana, J. M. (2012). A Sensor System for Automatic Detection of Food Intake Through Non-Invasive Monitoring of Chewing. *IEEE Sensors Journal, 12*(5), 1340-1348.

[2] Sazonov, E. S., Makeyev, O., Schuckers, S., Lopez-Meyer, P., Melanson, E. L., & Neuman, M. R. (2010, June). Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behaviror. *IEEE Transactions on Biomedical Engineering, 57*(3), 626-633.

[3] Stellar, E., & Shrager, E. (1985). Chews and swallows and the microstructure of eating. *The American Journal of Clinical Nutrition*, 973-982.

[4] Okazaki, H., Tsujimura, H., Doi, H., & Matsumura, M. (2009, October). Non-Restrictive Measurement of Swallowing Frequency by Throat Microphone. *IEICE Technical Committee Submission System, 109*(258), 1-4.

[5] Nagae, M., & Suzuki, K. (2011). A Neck Mounted Interface for Sensing the Swallowing Activity based on Swallowing Sound. *33rd Annual International Conference of the IEEE EMBS*, (pp. 5224-5227). Boston.

[6] Aboofazeli, M., & Moussavi, Z. (2004). Automated Classification of Swallowing and Breath Sounds. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, (pp. 3816-3819). San Francisco.

[7] Makeyev, O.; Sazonov, E.; Schuckers, S.; Lopez-Meyer, P.; Melanson, E.; Neuman, M. "Limited receptive area neural classifier for recognition of swallowing sounds using continuous wavelet transform", *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE,* On page(s): 3128 - 3131

[8] Loyd, L. (2007, 09 17). *Fighting Against Obesity One* . (P. Inquirer, Producer, & Philly.com) Retrieved 11 08, 2012, from Philly.com: http://www.smallbite.com/Docs/Fighting%20against%20obesity%20one%20bite%20at%20a%20time.pdf

[9] *What Are the Health Risks of Overweight and Obesity?* (2012, July 13). Retrieved October 31, 2012, from National Heart Lung and Blood Institute: http://www.nhlbi.nih.gov/health/health-topics/topics/obe/risks.html

[10] *Adult Obesity Facts*. (2012, August 13). Retrieved November 1, 2012, from Centers for Disease Control and Prevention: http://www.cdc.gov/obesity/data/adult.html

[11] *500m People Will Be Using Healthcare Mobile Applications by 2015*. (2010, November 10). Retrieved November 2, 2012, from Research2Guidance: Global Mobile Health Care Market Report 2010-2015: http://www.research2guidance.com/500m-people-will-be-using-healthcare-mobile-applications-in-2015

[12] Alan. (2011, November 10). *Wearable Device Captures Food Intake, Lifestyle Patterns*. Retrieved November 8, 2012, from Science Business: http://sciencebusiness.technewslit.com/?p=6895

[13] Glanz, K., Murphy , S., Moylan , J., Evensen , D., & Curb , J. (2006, Jan-Feb). Improving Dietary Self-Monitoring and Adherence with Hand-Held Computers: A Pilot Study. *American Journal of Health Promotion, 20*(3), 165-70.

[14] Ma, Y., Bertone, E., Stanek , E. 3., Reed, G., Hebert , J., Cohen , N., et al. (2003, July). Association Between Eating Patterns and Obesity in a Free-Living US Adult Population. . *Americal Journal of Epidemiology, 158*(1), 85-92.

[15] Ma, Y., Olendzki, B., Pagoto, S., Hurley, T., Magner, R., Ockene, I., et al. (2009, August). Number of 24-hour Diet Recalls Needed to Estimate Energy Intake. *Annals of Epidemiology., 19*(8), 553-9.

[16] Swartz, K. (2011, April 19). *Meal Snap - Calorie Counting Magic iPhone app doesn't dazzle*. Retrieved November 2, 2012, from Appolicious: http://www.appolicious.com/articles/7638-meal-snap-calorie-counting

[17] H. Kalantarian, S. I. Lee, A. Mishra, H. Ghasemzadeh, and M. Sarrafzadeh, "Multimodal energy expenditure calculation for pervasive health: A data fusion model using wearable sensors," *in Proceedings of IEEE PerCom Workshop on Smart Environments and Ambient Intelligence*. ACM, 2013

[18] H Kalantarian, N Alshurafa, M Sarrafzadeh, "A Wearable Nutrition Monitoring System," *IEEE Body Sensor Networks,* June 2014.