S2NI: A Mobile Platform for Nutrition Monitoring from Spoken Data

Niloofar Hezarjaribi[1], Cody A. Reynolds[1], Drew T. Miller[1], Naomi Chaytor[2], and Hassan Ghasemzadeh[1]
[1]Embedded & Pervasive Systems Lab (EPSL)
School of Electrical Engineering and Computer Science
Washington State University, Pullman, WA, 99164-2752
[2]Elson S. Floyd College of Medicine
Washington State University Spokane, Spokane, WA, 99210-1495
Email: {n.hezarjaribi, cody.reynolds, drew.t.miller, naomic, hassan.ghasemzadeh}@wsu.edu

Abstract— **Diet and physical activity are important lifestyle and behavioral factors in self-management and prevention of many chronic diseases. Mobile sensors such as accelerometers have been used in the past to objectively measure physical activity or detect eating time. Diet monitoring, however, still relies on self-recorded data by end users where individuals use mobile devices for recording nutrition intake by either entering text or taking images. Such approaches have shown low adherence in technology adoption and achieve only moderate accuracy. In this paper, we propose development and validation of *Speech-to-Nutrient-Information* (S2NI), a comprehensive nutrition monitoring system that combines speech processing, natural language processing, and text mining in a unified platform to extract nutrient information such as calorie intake from spoken data. After converting the voice data to text, we identify food name and portion size information within the text. We then develop a tiered matching algorithm to search the food name in our nutrition database and to accurately compute calorie intake. Due to its pervasive nature and ease of use, S2NI enables users to report their diet routine more frequently and at anytime through their smartphone. We evaluate S2NI using real data collected with 10 participants. Our experimental results show that S2NI achieves 80.6% accuracy in computing calorie intake.**

I. Introduction

The prevalence of chronic diseases and healthcare costs associated with conditions such as diabetes, cardiovascular disease, cancer, and obesity has been rising nationally and worldwide. In the United States, 30% of adults and 16% of youth suffer from obesity [1]. Moreover, medication costs associated with obesity are 9.1% of the United States' annual medical expenditures [2]. Cardiovascular disease has been the leading cause of death. Also, being overweight causes insulin resistance in diabetes patients. As a result, interventions that help with prevention or self-management of chronic diseases play a central role in reducing healthcare costs as well as mortality and morbidity rates.

Diet and physical activity are critical for self-management of many chronic diseases. The ability to objectively monitor physical activity and diet is of great importance in delivering timely and effective clinical interventions in both self-management and prevention of disease. Traditionally, self-reported approaches have been introduced for diet and physical activity monitoring [4]. The self-report approaches (e.g., questionnaires) have limitations such as low user compliance and lack of sufficient accuracy due to the subjective nature of monitoring [5], [6]. One of the self-monitoring techniques is the use of pen/pencil and paper diaries [7], [8]. Although these approaches have shown some success in weight loss programs [9], detailed self-monitoring can be cumbersome and time consuming; accordingly, adherence to these methodologies is low [10].

For this reason, a large group of researchers have focused on automating the task of monitoring physical activity and diet recently. Most proposed systems are based on wearable sensors [11], [12]. While objective and seamless monitoring of physical activity using wearable sensors has been a possibility and has been the focus of much research in the past [13]–[16], current technologies for nutrition monitoring still require involvement of the end users. Moreover, these technologies have been required the user to wear a specialized on-body sensor such as neck collar, microphone, and camera, which causes limited utilization of this technique.

There have been few research studies on using mobile phones for nutrition monitoring. Furthermore, a relatively large number of mobile apps for monitoring caloric intake and physical activity have been publicly available. In [19] an app, called Dietary Intake Monitoring Application (DIMA), has been designed for diet monitoring where several features including touch screen, visual interface barcode scanner, and voice recorder has been utilized. In [20], a prototype has been created to motivate healthy dietary by providing just-in-time messages using a barcode scanner. Yet, this prototype only sends motivational messages. Another method has been proposed in a study for taking photographs of the foods and transferring the data to lab for further analysis [21].

These technologies, using smart-phones, either require the user to enter the food intake information (e.g., food name, portion size, calorie intake) or require the user to take an image of what they eat [17], [18]. Those information are then searched for in a database and the amount of calorie intake is computed according to the suggested values in the database. These tools require the user to specify the timing of eating during data entry. In [22] the role of voice input has been studied in human-machine communication. It has been known that using voice input is more beneficial when users' hands and eyes are busy, keyboard of the device is small, user is disabled, and natural language interaction is preferred.

Our goal in this paper is to develop a framework for nutrition monitoring by utilizing speech recognition techniques in order to improve the ease of use of nutrition monitoring in different environments and make the process of reporting nutrition intake more automatic and accurate. To the best of our knowledge, the area of voice-based nutrition monitoring has not been explored by any other researchers to date.

Our contributions in this work are as follows. (1) We introduce a new framework for voice-based nutrition monitoring and propose a hierarchy of data processing modules including speech recognition, natural language processing, and text analysis in order to extract nutrient information from spoken language; (2) We develop 2-tier string search algorithms including exact matching and edit-distance-based approximate matching to search food items from a nutrition database and to compute nutrient information; (3) The system is evaluated with real data collected with 10 subjects in an experiment using that mimics noise-free as well as realistic noisy environments where the spoken data are entered into the system.

II. Speech to Nutrient Information

The proposed system aims to monitor food intake using speech recognition and natural language processing techniques. The user talks through a speech-to-text mobile application. The audio signal is converted to text. The text is processed in real-time and after finding the food name and portion size, it computes the amount of calories. In Figure 1, a block diagram of the system is shown. In what follows, each module is explained in detail.

A. Speech Recognition Module

Speech Recognition is a branch of pattern recognition, where input to the system is a stream of sampled speech data and a sequence of words is the desired output. The audio signal is matched against existing patterns which represent different sounds. There are four steps for converting audio to text. These steps are as follows:

1) **Sampling:** First step in speech to text transition is sampling the audio signal; in order to do this the speech is sampled to generate discrete signal and then by digitizing that signal, samples are generated.
2) **Endpointing:** In this step, the presence of the speech is differentiated from the non-speech regions.
3) **Feature Extraction:** Pattern matching on the raw sample streams is not efficient; therefore, important features must be extracted from the audio wave. The features are frequency domain attributes of the signal. Each phoneme has different frequency features called formants.
4) **Template matching:** The system is trained on a set of templates such as ”Yesterday”, ”Tomorrow ”. After speaking a word, the system attempts to match the word with the trained templates and outputs the most similar word.

B. Natural Language Processing Module

Natural Language Processing (NLP) is the ability of a machine to understand humans pronouncement as it is spoken in order to provide responses for them. Unlike the prior NLP implementation which was based on direct hand-coding of large rule sets, recent NLP algorithms are mostly based on Machine Learning (ML) techniques. In order to use ML techniques large corpora, which is a hand-annotated set of documents with correct values, should be analyzed. Different ML classes are used for NLP (e.g. decision trees, perceptron, Hidden Markov Model (HMM), etc.). In this paper, NLP is used for Part Of Speech (POS) tagging which is a supervised learning problem. In this paper, NLP is utilized for tagging sentences so that these tags can be processed later for finding nutrition specific data.

C. String Matching Module

After tagging the word using NLP, S2NI uses two methods for matching the detected words with the nutrition database; 1) Exact matching, 2) Approximate matching. Exact matching tries to find the food name which is exactly the same as the detected word in the sentence. Approximate matching has two steps for finding the food name; 1) first it searches for similar words in the database using edit distance algorithm, 2) it calculates the similarity probability of the two words and compares it with a predefined threshold.

D. Nutrition Database

For calculating calorie intake, a database of food names and their specified nutrient per serving has to be available for the system. For this purpose, United States Department of Agriculture (USDA) national database is used. The database is implemented as a dictionary, where food names are keys and nutrient values and portion sizes are the tags. The dictionary is a hash table with the time complexity for accessing the data being of O(1).

III. Validation

In this section the accuracy of the speech recognition engine and NLP module is assessed. For converting speech to text, Google's Voice recognition engine is utilized. The mobile app is modified in order to save the output in a text file for further analysis. The text file is tagged using POS implemented in Python. For assessing the system 4 different environments are used; no noise, street sound, music, and movie. A script is provided for the subjects containing 50 different food names. In order to increase the calorie calculation accuracy a similarity algorithm is embedded in the system. The algorithm tries to fix error of the app and compensate for incompleteness of database. Besides finding the most similar food name to the word that user expressed, the probability of similarity is used as a threshold for choosing the food name from the database. The accuracy values are calculated for 8 different thresholds. Our results include performance of the S2T application with and without presence of noise, and the performance of the calorie calculation using different threshold values.

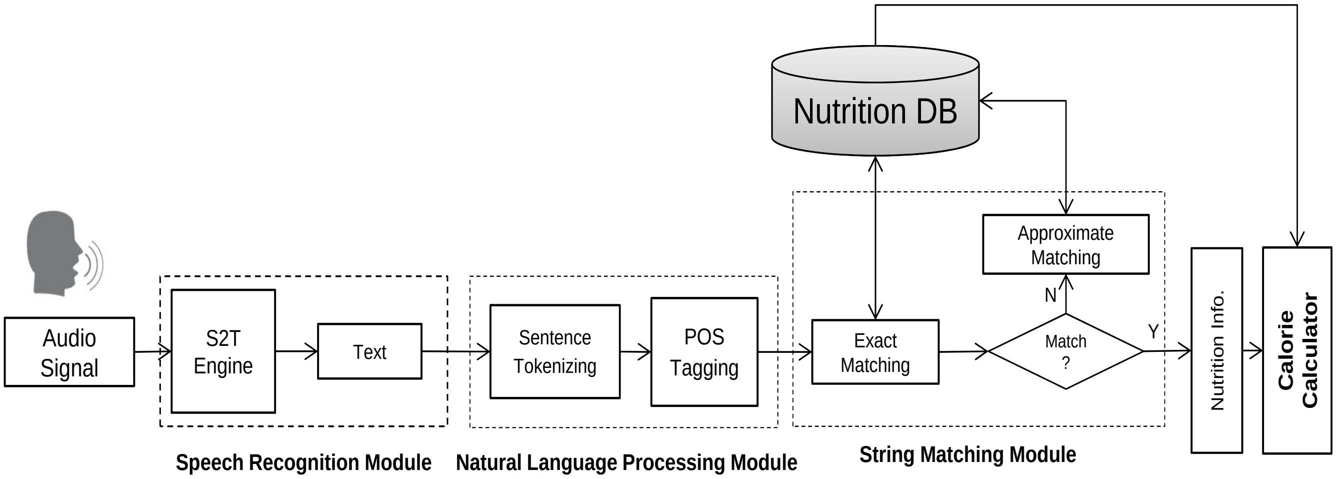

Fig. 1: Graphical illustration of the proposed S2NI system; including Speech Recognition Module, Natural Language Processing, String Matching Module, and Calorie intake computation

A. *Performance of Speech-to-Text (S2T) Module*

As mentioned previously, a script is provided to the subjects to read which is converted to text via the S2T app. Given that the focus of S2NI system is nutrition monitoring, the performance of the speech recognition module is computed using nutrition specific data. Further, the performance of the system is calculated using (1). TotalNutData is the total number of nutrition specific data in the script while TotalIncorrNutData is the total number of incorrectly detected data using the speech recognition module. The results are shown in Table I.

$$ACC_{S2Tapp} = \left(\frac{TotalNutData - TotalIncorrNutData}{TotalNutData}\right) \times 100 \quad (1)$$

TABLE I: Performance of pure output of S2T application

clearEnv	Street	Music	Movie
73.86%	63.26%	64.39%	64.39

B. *Performance of NLP Module*

After detecting the food names and their portion size, calorie value is calculated for the script. Since S2T application is not fully accurate, some of the food names are either not detected or incorrect. Moreover, the approximate matching tries to correlate non-nutrition specific data with the food names in the database. A threshold value is considered to address this issue. This value represents the probability of two words being similar. The program is tested with different threshold values. The results are shown in Table II. The first column is the performance of calorie calculation for the original script without utilizing S2T application. Other columns represent the performance of S2NI system after utilizing S2T application in presence of different background noises. The lower thresholds resulted in non-related data matching. In the other hand, by increasing the threshold beyond an optimal value, some of the nutrition specific data were not detected anymore (eg. ”Apple”, ”Apples”). We experimentally find the optimal value of the threshold. The results demonstrate that optimal threshold value is around 0.80. Performance and non-related data matching are the trade-offs of increasing and decreasing the threshold value away from 0.80, respectively.

TABLE II: Accuracy of the calorie calculation using the NLP module

threshold	script	clearEnv	Street	Music	Movie
0.70	97.16	92.03	86.27	79.64	87.58
0.74	96.31	86.43	73.41	81.68	81.84
0.78	96.03	88.30	76.45	85.81	80.96
0.82	96.03	89.36	75.34	85.32	81.31
0.86	96.03	87.44	75.27	85.32	81.80
0.90	96.03	85.31	72.10	84.08	81.80
0.94	90.86	81.22	58.90	62.26	72.42
0.98	87.69	77.83	58.20	60.96	70.37

C. *Performance of String Matching Module*

NLP module contains two matching terminologies; 1) Exact matching, 2) Approximation matching. In exact matching, the algorithm checks for the word detected as food name in a dictionary and finds the name that exactly matches with the detected word. If the word is not found in the dictionary, it tries to find the most similar food name to that. Approximation matching is implemented for improving the performance of speech recognition module. Figure 2 shows the performance of S2T application with and without presence of approximation matching. The results demonstrate that amount of improvement in finding Nutrition Specific data using approximate matching is 7%.

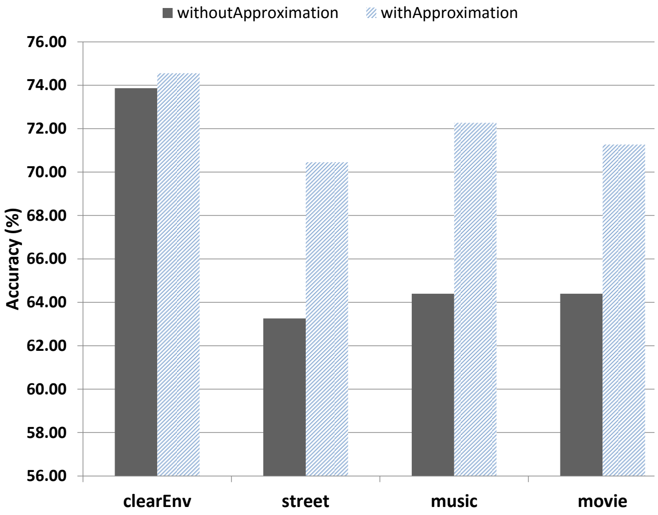

Fig. 2: Performance of finding nutrition-specific data with and without presence of approximate matching algorithm

IV. Conclusion and Ongoing Research

The major innovation of this paper is the introduction of an entirely novel approach for nutrition monitoring. This research integrates advances in speech recognition, natural language processing, text analysis, and mobile health in order to provide a more pervasive approach for recording and understanding spoken language for diet assessment. In using speech-to-text app, our goal was to provide a more convenient tool for users in different situations to record their food intake data. We also utilized natural language processing algorithms to identify nutrition-specific information within the generated text. Furthermore, we devised a 2-layer approach for analyzing the identified text to compute calorie intake. The results show that the performance of the S2T app is 66.50%. For extracting the nutrition specific data from the S2T output, NLP program was implemented in Python. The performance of the string matching module in presence of error-free text is 95%. When the matching algorithm takes the output of the S2T app as input, we still achieve 80% accuracy in computing the calorie intake information.

Our ongoing evaluation of S2NI involves in-the-wild testing where participants use the platform in their natural setting and beyond limitation of a laboratory-based experiment to record their food intake data and evaluate the usability and acceptability of the devised mobile health technology. For further improving the performance of the S2NI framework, we are currently working on expanding the capabilities of our S2T mobile app to recognize the spoken language of non-native speakers. Lastly, we are developing a prompting interface to allow for end users to not only enter food items that are not included in our nutrition database but also correct any misclassified or mis-identified food item.

References

[1] R. W. Kimokoti and B. E. Millen, “Diet, the global obesity epidemic, and prevention,” *Journal of the American Dietetic Association*, vol. 111, no. 8, pp. 1137–1140, 2011.

[2] E. A. Finkelstein, I. C. Fiebelkorn, G. Wang *et al.*, “National medical spending attributable to overweight and obesity: how much, and who’s paying?” *Health affairs-millwood va then bethesda ma*, vol. 22, no. 3; SUPP, pp. W3–219, 2003.

[3] M. Nestle, *Food politics: How the food industry influences nutrition and health*. Univ of California Press, 2013, vol. 3.

[4] L. R. Wilkens and J. Lee, *Nutritional epidemiology*. Wiley Online Library, 1998.

[5] K. B. Michels, “A renaissance for measurement error,” *International journal of epidemiology*, vol. 30, no. 3, pp. 421–422, 2001.

[6] D. R. Jacobs Jr, “Challenges in research in nutritional epidemiology,” in *Nutritional Health*. Springer, 2012, pp. 29–42.

[7] L. E. Burke, S. M. Sereika, E. Music, M. Warziski, M. A. Styn, and A. Stone, “Using instrumented paper diaries to document self-monitoring patterns in weight loss,” *Contemporary clinical trials*, vol. 29, no. 2, pp. 182–193, 2008.

[8] R. L. Collins, T. B. Kashdan, and G. Gollnisch, “The feasibility of using cellular phones to collect ecological momentary assessment data: application to alcohol consumption.” *Experimental and clinical psychopharmacology*, vol. 11, no. 1, p. 73, 2003.

[9] M. J. Devlin, “Obesity: Theory and therapy,” *The Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 6, no. 3, pp. 325–327, 1994.

[10] M. K. Mattfeldt-Beman, S. A. Corrigan, V. J. Stevens, C. P. Sugars, A. T DALCIN, M. J. Givi, and K. C. Copeland, “Participants evaluation of a weight-loss program,” *Journal of the American Dietetic Association*, vol. 99, no. 1, pp. 66–71, 1999.

[11] R. Saeedi, N. Amini, and H. Ghasemzadeh, “Patient-centric on-body sensor localization in smart health systems,” in *Signals, Systems and Computers, 2014 48th Asilomar Conference on*. IEEE, 2014, pp. 2081–2085.

[12] N. Hezarjaribi, R. Fallahzadeh, and H. Ghasemzadeh, “A machine learning approach for medication adherence monitoring using body-worn sensors,” in *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2016, pp. 842–845.

[13] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, “A wearable nutrition monitoring system,” in *Wearable and Implantable Body Sensor Networks (BSN), 2014 11th International Conference on*. IEEE, 2014, pp. 75–80.

[14] E. Thomaz, I. Essa, and G. D. Abowd, “A practical approach for recognizing eating moments with wrist-mounted inertial sensing,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1029–1040.

[15] P. Alinia, R. Saeedi, R. Fallahzadeh, A. Rokni, and H. Ghasemzadeh, “A reliable and reconfigurable signal processing framework for estimation of metabolic equivalent of task in wearable sensors,” *IEEE Journal of Selected Topics in Signal Processing*, 2016.

[16] H. Ghasemzadeh, R. Fallahzadeh, and R. Jafari, “A hardware-assisted energy-efficient processing model for activity recognition using wearables,” *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 21, 2016.

[17] C. Gurrin, Z. Qiu, M. Hughes, N. Caprani, A. R. Doherty, S. E. Hodges, and A. F. Smeaton, “The smartphone as a platform for wearable cameras in health research,” *American journal of preventive medicine*, vol. 44, no. 3, pp. 308–313, 2013.

[18] R. Almaghrabi, G. Villalobos, P. Pouladzadeh, and S. Shirmohammadi, “A novel method for measuring nutrition intake based on food image,” in *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*. IEEE, 2012, pp. 366–370.

[19] K. Connelly, K. A. Siek, B. Chaudry, J. Jones, K. Astroth, and J. L. Welch, “An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability,” *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 705–712, 2012.

[20] S. S. Intille, C. Kukla, R. Farzanfar, and W. Bakr, “Just-in-time technology to encourage incremental, dietary behavior change,” in *AMIA Annual Symposium Proceedings*, vol. 2003. American Medical Informatics Association, 2003, p. 874.

[21] C. K. Martin, H. Han, S. M. Coulon, H. R. Allen, C. M. Champagne, and S. D. Anton, “A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method,” *British Journal of Nutrition*, vol. 101, 2 2009.

[22] P. R. Cohen and S. L. Oviatt, “The role of voice input for human-machine communication,” *proceedings of the National Academy of Sciences*, vol. 92, no. 22, pp. 9921–9927, 1995.