

Pervasive Eating Habits Monitoring and Recognition through a Wearable Acoustic Sensor*

Yin Bi
Northeastern University, China
biyin0125@126.com

Wenyao Xu
The State University of New
York at Buffalo
wenyaoxu@buffalo.edu

Nan Guan
Northeastern University, China
guannan@ise.neu.edu.cn

Yangjie Wei
Northeastern University, China
weiyangjie@ise.neu.edu.cn

Wang Yi
Northeastern University, China
wangyi@ise.neu.edu.cn

ABSTRACT

Eating habits provide clinical diagnosis evidences of lifestyle related diseases, such as dysphagia and indigestion. However, it is costly to obtain eating habit information of common people in terms of both time and expenses. This paper presents a pervasive approach for eating habit monitoring and recognition by a necklace-like device and a smartphone communicating via bluetooth. The necklace-like device acquires acoustic signals from the throat, and the data are processed in the smartphone to recognize important features. With complex acoustic signals collected from the throat, our method comprehensively analyzes and recognizes different events including chewing, swallowing, and breathing in the smartphone. Experiments show that the proposed approach can recognize different acoustic events effectively, and the recognition accuracy with K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) is 86.82% and 98.35%, respectively. Finally, a real eating case study is conducted to validate the proposed approach.

Keywords: Eating habit, Feature extraction, KNN, SVM

1. INTRODUCTION

Food selection and food intake patterns are vital elements to describe an individual's eating habits, which can be defined as what and how people eat in their daily lives. Eating habits monitoring is a prerequisite for effective diagnosis, prevention, and intervention [5]. In recent years, as people's concerns of health increase rapidly, eating habit monitoring becomes a hot research topic of doctors, nutritionist and researchers in many other fields. In addition to what type of food we can sense and manage, how people eat, including the way of chewing and swallowing during eating, is even more important for health evaluation and management [18,

*This work is partially supported by NSF of China under grant no. 61300022 and 61305025.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PervasiveHealth '14 May 20–23, 2014, Oldenburg, Germany
Copyright 2014 ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

10]. Eating habits provide important clinical evidences for diseases like dysphagia [12], indigestion [5].

Traditionally, eating habits monitoring is conducted in clinical environments, which is inefficient in both time and expenses. In this paper, we use a wearable acoustic sensor system to continuously acquire real-time eating habit information of common people in an economical, convenient and non-interfering way. Existing wearable acoustic sensor systems focus on the detection of one particular sound [18, 10, 13, 9, 16], which is inadequate for precise eating habit monitoring and recognition.

In this paper, we propose a pervasive solution to detect and analyze complex acoustic signals during eating. The acoustic signals during eating are collected by a necklace-like device directly from the throat. We propose a comprehensive solution to detect different acoustic events, and extract features in the time and frequency domain for each event. We sort the feature with Relief-F to create different feature subsets. Finally, we use both KNN and SVM to recognize acoustical events with respect to different feature subsets.

1.1 Related Work

Individual sensing provides important information of eating habits [2]. In particular, measuring vibrations during food chewing, swallowing, breathing and speaking is vital sensing for automatic health monitoring systems. A work recording with acoustic sensors during chewing crisp and hard foods were performed by Drake [8]. Subsequent studies focused on relating chewing sounds to sensory perception of foods and to food assessments [22]. Attempts to classify foods using pattern recognition techniques were performed by De Belie et al. [7]. Their work showed that a small number of foods can be classified in a laboratory setting using foam-based ear sensors. In addition, they found that chewing sound patterns changed during the breakdown process of several chewing cycles. This observation was later confirmed using an automatic unsupervised sequence searching technique to group chewing strokes [3]. Lopez-Meyer et al. used microphones to record chewing and swallowing sounds and detect periods of food intake by evaluation of the instantaneous swallowing frequency [17]. Lee et al. [15] investigated the chewing sounds recorded outside the closed mouth of the participant, who is chewing potato chips and tortilla chips. They discovered that the energy of the spectrum of a chewing event decreases in the progress of food crushing. In [1], researchers established the viability of a reduced occlusion

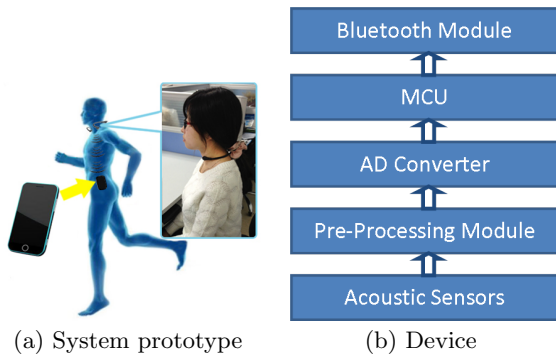


Figure 1: Illustration of the Hardware System

sensor prototype to detect chewing sounds.

However, the proceeding concept was based on chewing and swallowing sounds recording and counting, and no virtually data processing algorithm has been proposed. Moreover, during chewing and swallowing, other sounds such as breathing and speaking can also be recorded at the same time. In fact, most of time, the record is the mixture of all these sounds, so the identification and classification is essential for a more accurate monitoring algorithm.

2. HARDWARE SYSTEM

The eating habit monitoring system consists of a necklace-like device and a smartphone, which communicate with each other via bluetooth, as shown in Figure 1-(a). The necklace-like device samples acoustic signals from the user's throat, and upload data to the smartphone. An APP on the smartphone executes the recognition algorithms. The necklace-like device consists of five main parts: a pair of acoustic sensors, a pre-processing module, an AD converter, an ultralow-power MCU and a communication module, as shown in Figure 1-(b). The acoustic sensors are placed near the user's throat, to precisely collect the swallowing, breathing and chewing signals. The pre-processing module amplifies and filters input acoustic signals. The first-stage amplified circuit is connected to the sensor directly and its gain is 5. The second stage normal-phase amplifier uses LM358, which has the advantage of high common-mode rejection ratio, low noise and high gain. In addition, the total gain of the amplifier is 250 and the cutoff frequency of the low-pass filter is 3000 Hz. The AD converter TLV2541 digitizes the analog signals with a sampling rate of 220HZ. The ultralow-power MCU MSP430F5438, widely used in energy-constrained consumer electronics products and portable medical equipments, is responsible for the control and data transmission via the bluetooth communication module.

3. RECOGNITION

This section introduces the recognition framework, which consists of acoustic events detection, acoustic events recognition and eating-habit index value computation.

3.1 Detection

Swallowing, chewing and breathing samples are detected from received sequential data. We assume that all information of each event can be extracted from each single window with a fixed length of 100. After normalizing the sequential data, once a value of the normalized data exceeds the

pre-defined threshold value, an acoustic event will appear in the following sequential data, so from this point to the following 99 points, they are combined to an acoustic event. In this way, we extracted each sample contains 100 points, and every sample only contains one unabridged event.

3.2 Recognition

Based on the detected events in the received sequential data, the features of each sample are extracted. In this paper, we mainly focus on statistical features in the time and frequency domain. The following shows the list of our statistical features in time domain, and most of them have been intensively investigated in the previous studies and proved to be useful in the recognition problems [4, 19, 14]: (1) Max: Maximum value of a sample; (2) Min: Minimum value of a sample; (3) Mean: Average value of a sample; (4) Variance: The square of Std_variance; (5) Std_variance: Measure of the spreadness of the sample over the window; (6) Sample entropy: Measure of the complexity of the sample; (7) Skewness: The degree of asymmetry of the sample distribution; (8) Kurtosis: The degree of peakedness of the sample distribution; (9) Interquartile: Measure of the statistical dispersion, range being equal to the difference between the 75th and the 25th percentiles of the sample over the window; (10) Zero crossing: The total number of times the sample rate changes from positive to negative or the other way around, normalized by the window length.

In the frequency domain, Fast Fourier transform (FFT) is first used to extract the signal features, and they are estimated from Power Spectral Density (PSD) based on Welch's method with a Hamming window. The length of the Hamming window is 40 and the overlap adjacent windows is 25% [23]. The maximum power (Pmax), average power (Pmean) of the sample, power in 0-15Hz, 15-30Hz, 30-45Hz, 45-100Hz, 100-200Hz, 200-300Hz, 300-400Hz and 400-512Hz are calculated with respect to PSD of the sample.

The next step in recognition is to sort the features and select the most important features for different acoustic events discrimination. Because the scale factors and units of the proceeding features are different, we normalize all the features to zero mean and unit variance using:

$$f_{normalized} = \frac{f_{raw} - \mu}{\sigma} \quad (1)$$

where μ and σ are the empirical mean and the standard deviation of a particular feature across all activity classes.

Then, Relief-F is used for our feature selection due to its high popularity and usefulness in many pattern recognition and machine learning problems. Relief-F is a popular filter method that estimates the relevance of features according to how well their values distinguish between the data points of the same and different classes that are near each other [24]. Specifically, it computes a weight for each feature to quantify its merit. This weight is updated for each of the data points presented, according to the evaluation function:

$$w_i = \sum_{j=1}^N \left(x_i^j - nearmiss(x^j)_i \right)^2 - \left(x_i^j - nearhit(x^j)_i \right)^2 \quad (2)$$

where w_i is the weight of the i^{th} feature, x_i^j is the value of the feature for data point x^j , N is the total number of the data points, $nearmiss(x^j)$ and $nearhit(x^j)$ are the nearest points to x^j from the same and different class, respectively.

The higher is the weight, the more important is the feature. In order to recognize these events including swallowing, breathing and chewing, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are used in this paper, both of which are widely used in recognition problems.

KNN [6] is one of the oldest and simplest methods for pattern recognition. The basic principle of KNN is that samples residing closer in the instance space have the same class values, and the k-nearest neighbor classification takes k nearest samples of the testing sample to make a decision.

When a new sample x arrives, KNN finds the k neighbors nearest to the unlabeled sample from the training space based on certain distance measurements. In this work, the Euclidean distance is used. First, we use $N_k(x)$ to denote the k labeled samples nearest to x and $c(z)$ the class label of z . Then the subset of the nearest neighbors within class $j \in \{1, 2, \dots, l\}$ (l is the number of classes) is

$$N_k^j = \{y \in N_k(x) : c(y) = j\} \quad (3)$$

Finally, the classification result $j \in \{1, 2, \dots, l\}$ is defined as a majority vote:

$$j^* = \operatorname{argmax}_{j=1,2,\dots,l} |N_k^j(x)| \quad (4)$$

Nowadays, SVM [20], as the best separating (maximal margin) hyperplane between the two classes of training samples in the feature space, has become one of the most successful approaches in pattern recognition. In this work, we choose SVM with a linear kernel $f(x) = \langle w, x \rangle + c$ as a learning machine, which corresponds to a hyperplane. For a given pattern mapped in the feature space to x , the value of $f(x)$ is a positive number, then the pattern belongs to the class labeled by the numeric value +1; otherwise, it belongs to the class with value -1. In order to recognize multi-class samples, we choose one-versus-one (OVO) [11], which designs SVM between any two types of samples, and K types of samples requires $k(k-1)/2$ SVM. When an unknown sample is to be recognized, the final classification of this sample is the votes of these SVM.

3.3 Indexing

After recognize the acoustic events, we need to calculate the eating-habit index value to quantitatively evaluate the eating habits. In this paper, the total number of swallowing, the average number of swallowing per unit time, the average time of chewing before a swallowing and the average number of chewing before a swallowing are used as index values.

4. EXPERIMENTS

4.1 Data Collection

In the recognition experiment, swallowing, chewing and breathing signals are recorded from a healthy male subject, seated in a relatively quiet laboratory. The subject is required to swallowing water, chewing potato chips and breathing as he usually does. No-event samples are recorded when the throat microphone is hung in the air. After it is worn by the subject, we collect 30 breathing, 30 chewing, 30 swallowing and 30 No-event samples. In the indexing and habit discovery experiment, the subject seats in a quiet laboratory, and eats potato chips in single pieces. Then he is asked to chew and swallow without speaking. The actual number of chewing and swallowing when he eats is counted.

4.2 Event Recognition

After collecting 30 breathing, 30 chewing, 30 swallowing samples and 30 No-event events, features in the time and frequency domain are extracted for each event. Then, 20 features are sorted with Relief-F, where different feature subsets are created with the first k ordered feature, and

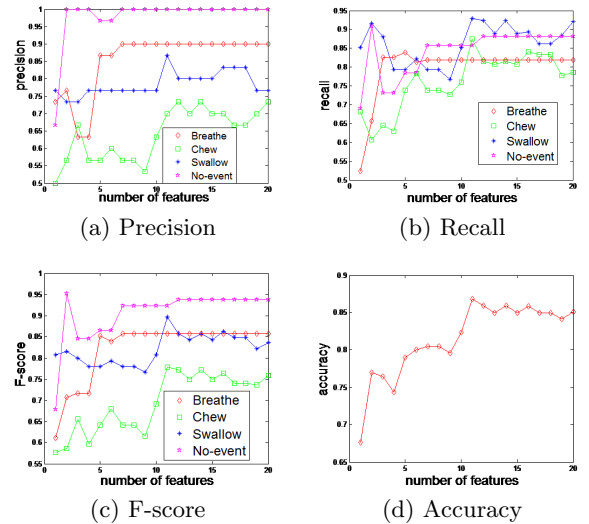


Figure 2: Results of recognition by KNN for different feature subsets from each feature group

k changes from 1 to the total number of extracted features with a step-size of 1. Subsequently, KNN and SVM are used to classify these three activities based on the sorted feature subsets, respectively, the results of which are shown in Figure 2 and Figure 3. Since the samples are limited, these 90 samples are used as the training set and testing set. The recognition performance is evaluated with:

$$Precision = TP / (TP + TN)$$

$$Recall = TP / (TP + FN)$$

$$Accuracy = (TP + FN) / (TP + TN + FP + FN)$$

$$F\text{-score} = 2 \times Precision \times Recall / (Precision + Recall)$$

where TP, TN, FP and FN represents the number of true positive, true negative, false positive and false negative classified samples, respectively [21]. Figure 2 shows the classification performance by KNN when K is 11. Each line in (a), (b) and (c) represents different acoustic events. With respect to the precision, recall, F-score and accuracy results of different feature subsets, it can be seen that using feature subsets with 11 features gives the optimum number of features for breathing, chewing, swallowing and no-event recognition. The recognition accuracy is 86.82%. Furthermore, the recognition precision of breathing and no-event is much higher than that of chewing and swallowing, and some of chewing samples is recognized as breathing samples falsely. If we pick more features beyond the top 11, the recognition performance only varies slightly.

Figure 3 shows the classification performance of SVM. It can be seen that using feature subsets with 17-19 features gives the optimum number of features for breathing, chewing, swallowing and No-event recognition, with a total accuracy of 98.35%. Furthermore, when the number of features increases from 1 to 11, the recognition performance improves rapidly. If we pick more features beyond the top 10, the recognition accuracy is more than 95%.

From above results, we can see that in order to recognize breathing, chewing, swallowing and No-event samples more precisely, classifier SVM has better recognition performance than classifier KNN. So in the following, we use SVM with 20 feature subsets to recognize acoustic events.

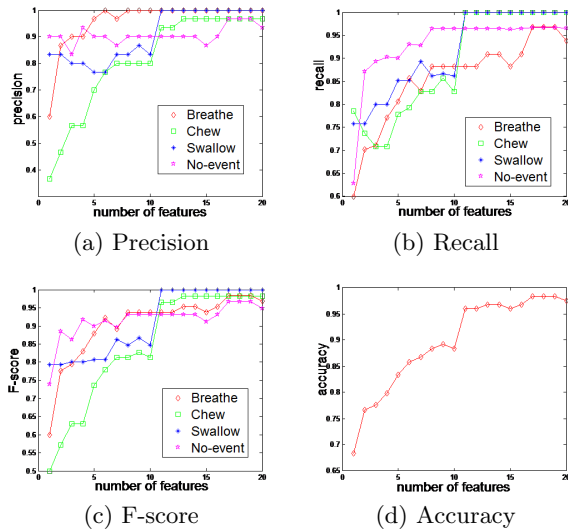


Figure 3: Results of recognition by SVM for different feature subsets from each feature group

Table 1: Eating habit index of the case study

Index	Index Value
Ratio between chewing and swallowing	5.13
Average number of swallowing in one minute	2.53
Average time of chewing	0.26 (second)
The longest time of chewing	0.41 (second)
The shortest time of chewing	0.12 (second)
Average number of chewing before a swallowing	5.21
The largest number of chewing before a swallowing	34
The least number of chewing before a swallowing	0

4.3 Indexing and Habit Discovery Case Study

We continuously collect the sampling data of eating potato chips for about 15 minutes, and detect 398 acoustic events in the collected data. Then we extract features in the time and frequency domain for each event and the classifier SVM is used to recognize these events. In the 398 acoustic events, we recognize 195 chewing events and 38 swallowing events, and the chewing time is calculated according to the feature of zero crossing rate. The eating habit indexing value of this real case study is shown in Table 4.3, from which we can quantitatively evaluate the eating habit of the subject.

5. CONCLUSION

Eating habits related diseases are growing problems of the world's population, and pervasive eating habit monitoring provides important benefit to recognize these diseases in advance. In this paper, a comprehensive approach using a necklace-like device is presented. First, in order to systematically assess the usefulness and identify the most important features, Relief-F is used to select the needed features. Then in the recognition process, KNN and SVM are both introduced to recognize the acoustic events, and the recog-

nition accuracy of the acoustic events is 86.82%, 98.35%, respectively. Finally, a real eating case is monitored and recognized to validate our scheme. Our approach not only can recognize the acoustic events, but also calculate eating habit index values to quantitatively evaluate his eating habits. In the future work, we will adopt filter algorithms to improve the precision of our monitoring algorithm.

6. REFERENCES

- [1] O. Amft. A wearable earpad sensor for chewing monitoring. *IEEE Sensor Conference*, 2010.
- [2] O. Amft. Ambient, on-body, and implantable monitoring technologies to assess dietary behaviour. *International Handbook of Behavior, Diet and Nutrition*, Springer, 2011.
- [3] O. Amft, M. Kusserow, and G. Troster. Automatic identification of temporal sequences in chewing sounds. *BIBM*, 2010.
- [4] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive Comput.*, 2004.
- [5] F. Bellisle. Why should we study human food intake behaviour? *Nutr Metab Cardiovasc Dis.*, 13(4), 2003.
- [6] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. on Information Theory*, 1967.
- [7] N. DeBelie and V. D. Smedt. Principal component analysis of chewing sounds to detect differences in apple crispness. *Postharvest Biol Technol*, 18, 2000.
- [8] B. Drake. Food crushing sounds. an introductory study. *J Food Sci*, 28(2), 1963.
- [9] K. O. et al. No contact-type chewing number counting equipment using infrared sensor. *T. SICE*, 38(9), 2002.
- [10] M. S. et al. Wearable eating habit sensing system using internal body sound. *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 4(1), 2010.
- [11] R. Z. et al. Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2007.
- [12] A. Farri, A. Accornero, and C. Burdese. Social importance of dysphagia: its impact on diagnosis and therapy. *ACTA Otorhinolaryngo Italica*, 2007.
- [13] M. Furuta and H. Miyako. Observation of matters related to chewing habits. *Japanese Journal of Psychosomatic Dentistry*, 14(2), 1999.
- [14] T. Huynh and B. Schiele. Analyzing features for activity recognition. *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence*, 2005.
- [15] W. L. III, A. Deibel, C. Glembin, and E. Munday. Analysis of food crushing sound during mastication: Frequency-time studies. *Journal of Texture Studies*, 19(1), 1988.
- [16] G. Lopez and I. Yamada. New healthcare society supported by wearable sensors and information mapping based services. *WIVE*, 2009.
- [17] P. Lopez-Meyer, S. Schuckers, O. Makeyev, and E. Sazonov. Detection of periods of food intake using support vector machines. *EMBS*, 2010.
- [18] J. Nishimura and T. Kuroda. Eating habits monitoring using wireless wearable in-ear microphone. *ISWPC*, 2008.
- [19] N. Ravi, N. Dandekar, and P. M. et al. Activity recognition from accelerometer data. *AAAI*, 2005.
- [20] V. Vapnik. *Statistical learning theory*. 1998.
- [21] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. *IJCAI*, 1999.
- [22] Z. M. Vickers. Food sounds: How much information do they contain? *J Food Sci*, 45(6), 1980.
- [23] A. Yadollahi and Z. Moussavi. Feature selection for swallowing sounds classification. *EMBC*, 2007.
- [24] M. Zhang and A. Sawchuk. A feature selection-based framework for human activity recognition using wearable multimodal sensors. *BODYNETS*, 2011.