An Audio-based Hierarchical Smoking Behavior Detection System Based on A Smart Neckband Platform

Jinqi Cui, Liang Wang*
State Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
jinqicui@gmail.com
wl@nju.edu.cn

Tao Gu
School of Computer Science and Information Technology
RMIT University, Australia
tao.gu@rmit.edu.au

Xianping Tao, Jian Lu
State Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
{txp,lj}@nju.edu.cn

ABSTRACT

Smoking behavior detection has attracted much research interest for its significant impact on smokers' physical and mental health. Existing research has shown the potential of using wearable devices for fine-grained smoking puff and session detection by detecting a smoker's content of breathing, lighter usage, breathing, and gesture patterns. However, the existing systems are complex, and they are usually vulnerable to confounding activities and diversity of smoking behavior. To address these limitations, this paper proposes the design and implementation of a simple and compact smart neckband device for smoking detection. The device is equipped with both passive and active acoustic sensors to detect smoking sessions and puffs. We propose a hierarchical processing framework in which the lower-layer detects the sub-movements, i.e., lighter usage, hand-to-mouth gesture and deep breathing, from perceived audio data; and the higher-layer, based on the lower-layerąŕs detection results, detects smoking puffs and sessions using temporal sequence analysis techniques. Real-world experiments suggest our system can accurately detect smoking puffs and sessions with F_1 *score* of respectively 93.59% and 92.96% in complex environments with the presence of confounding activities and diverse ways of smoking.

CCS Concepts

•**Human-centered computing** → Ubiquitous and mobile computing systems and tools;

Keywords

Smoking Detection; Wearable Acoustic Sensing; Hierarchical; Temporal Sequence Processing;

*Corresponding author

MOBIQUITOUS '16, November 28-December 01, 2016, Hiroshima, Japan
© 2016 ACM. ISBN 978-1-4503-4750-1/16/11...$15.00
DOI: http://dx.doi.org/10.1145/2994374.2994384

1. INTRODUCTION

According to the National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health, cigarette smoking causes more than 480,000 deaths each year in the United States [7]. This is nearly one in five deaths. Besides its significant impact on human physical health, cigarette smoking has also known to be correlated to the smoker's mental state. Regular smokers report negative moods when they have not recently smoked [14]. It has also been discovered that people with difficulty to regulate their mood are more vulnerable to drugs like nicotine [8]. As a result, monitoring a smoker's smoking behavior gives us an opportunity to understand the smoker's physical and mental health. And we can further develop applications that keep smokers aware of their amount of intake, motivate them to quit smoking, or detect depression by detecting tobacco abusement.

In this paper, we focus on the problem of detecting detailed daily smoking behavior using wearable sensors. More specifically, we aim at detecting not only the number of cigarettes smoked but also the number of smoking puffs taken during each smoking session. Existing smoking behavior detection approaches mainly fall into two categories—self-reporting and automatic detection. The self-reporting approach asks users to fill in questionnaires regarding their daily smoking behavior [21]. By relying on human memory, this approach is not affected by confounding events and behavior diversity, and can collect data such as the smoker's mental state [8]. However, self-reporting requires extensive human effort and is potentially unreliable for high-risk populations [6]. As a result, there is a growing research interest in automatic smoking behavior detection using wearable sensors. Some work explores lighter usage as a signature for smoking [20]. However, they rely on a customized lighter and cannot track the smoking puffs within a smoking session. Deep breathing is explored by some researchers as a key indicator to identify smoking puffs [4, 13]. However, there are many other activities that involve deep breathing, such as exercising and yawning. Other work explores the unique gesture patterns of the smoking behavior for fine-grained smoking detection [19, 16]. However, similar to deep breathing-based approaches, gesture-based approaches are also vulnerable to confounding events with similar gestures like eating and drinking. Some existing medical studies [9, 22] have used specialized devices such as a CO monitor to record and analyze the smoking topography for smokers under designed experimental settings. However, their approach is not applicable to unconstrained daily scenarios for requiring specialized medical devices and human participation.

Despite the above efforts, smoking detection in unconstrained daily life scenarios remains a challenging issue mainly due to the following reasons. 1) **System Complexity**—the simple smoking behavior involves a series of movements including lighting the cigarette, taking the cigarette to the mouth and puff. Detecting these movements may require different devices attached to different body parts or objects [20, 4, 13, 19, 16], resulting a complex system difficult to wear and obtrusive; 2) **Confounding activities**—in an unconstrained daily life scenario, the smoking behavior is often similar to other behaviors like drinking, eating, yawning, etc; 3) **Behavior diversity**—the smoking behavior can be carried out in diverse ways: the smoker may hold the cigarette using the dominant or non-dominant hand; smoking may also be concurrent with other behaviors like talking, walking, eating, etc. The major limitation of existing approaches lies in that they require different devices to capture different movements, and only explore a single aspect of the smoking behavior—lighter usage [20], deep breathing [4, 13], or gesture [19, 16], which can easily be confused with other confounding activities and vulnerable to the diversity of smoking behavior.

In this work, we address the above challenges by proposing a simple and effective solution to the smoking behavior detection problem. First, to reduce system complexity, we propose a simple and compact smart neckband hardware platform equipped with both passive and active acoustic detection devices to capture physical data from smoking behaviors. Second, to discriminate smoking against other confounding activities, we propose a hierarchical data processing framework that first detects three key smoking-related sub-movements, i.e., lighter usage, hand-to-mouth gesture, and deep-breathing. Smoking puff and session detection is then performed based on the unique patterns of these sub-movements which are discriminative against other activities. We adopt Allen's interval algebra [5] to describe the subtle temporal structure of these key sub-movements composing each smoking puff, which has shown to be promising. Based on the accurate detection of smoking puffs, smoking session is done by applying temporal clustering over the detected smoking puffs. Finally, the proposed system handles the diversity of smoking behavior by two means: 1) the above hierarchical data processing framework can accurately detect smoking behavior even with other activities performed concurrently because it is based on a sophisticated model that takes multiple smoking-related sub-movements into account; 2) the smart neckband is so designed that it can accurately capture the smoking-related sub-movements, especially for the hand-to-mouth gesture, regardless the hand used or the posture of the smoker.

The proposed system is evaluated through extensive experiments using both short- and long-term data collected in various environments by several smokers. Experiment results suggest our system can detect smoking puffs and sessions accurately with F_1 *score* of 93.59% and 92.96%, respectively, in various environments with the presence of multiple confounding and concurrent activities.

In general, this paper makes the following contributions:

- We propose a sophisticated hierarchical approach that detects smoking puffs and sessions based on a subtle temporal model of the key sub-movements involved in smoking behaviors.
- We design and implement a simple and compact smart neckband equipped with passive and active acoustic detection abilities that can detect the diverse smoking-related sub-movements accurately.
- We conduct extensive experiments using data collected from multiple smokers in unconstrained real-life scenarios.

The rest of the paper is organized as follows. Sec. 2 summaries the related work. Sec. 3 presents an overview of the system. Detailed system design is presented in Sec. 4. Sec. 5 introduces the data collection and system evaluation results. Finally, Sec. 6 concludes the paper.

2. RELATED WORK

Existing medical studies [12, 9, 22] have used specialized medical devices such as CO monitors such as CReSS Pocket [1] or Micro+ [3] to record and analyze details of smoking behaviors. By smoking through a mouthpiece attached to the device, it records the user's smoking behavior (including puffs and timestamps) as well as the CO levels in a single breath. However, the specialized medical devices are expensive and complex, which are not feasible for unconstrained daily scenarios.

Detecting smoking-related movements is another possibility for smoking detection, such as detecting lighter usage, gesture of smoke and respiration of smoke. UbiLighter [20] tracks the time-of-day and number of consumed cigarettes of users using an instrumented rechargeable USB lighter. Though it is possible to count the number of lighter usage using UbiLighter, this approach does not provide information about detailed smoking puffs and its duration, which is particularly useful for determining the degree of nicotine intake across subjects [9, 10]. In [13], Sazonov et.al. presents a hand gesture tracking device using a miniature RF transmitter worn on the wrist and an antenna worn on the chest to detect the hand-to-mouth gesture performed during smoking. However, their system is difficult to wear and can easily confuse smoking with other gestures like eating and drinking. RisQ [16] proposes to use a wrist-worn inertial sensor to identify smoking-related gestures. However, their approach is still vulnerable to confounding events and can only track the movement of a single hand (dominant or non-dominant) while the smoker can typically use either hand to hold the cigarette. mPuff [4] is proposed to perform smoking detection by measuring smoker's respiration, and shows that a SVM classifier can identify smoking episodes on per-subject basis. Their approach requires the user to wear a respiratory inductive plethysmograph chest band which is cumbersome and uncomfortable. Also, tracking respiration along can make the system vulnerable to other confounding events such as taking a deep-breath or yawning. Smokey [24] provides a device-free smoking detection approach using commercial WiFi infrastructures. However, this approach is infrastructure dependent and also suffers from the limitations of gesture-based smoking detection methods. The proposed system in this paper differs from the above work in two major aspects: 1) the smart neckband we designed is compact and can reliably detect smoker's gesture regardless the hand

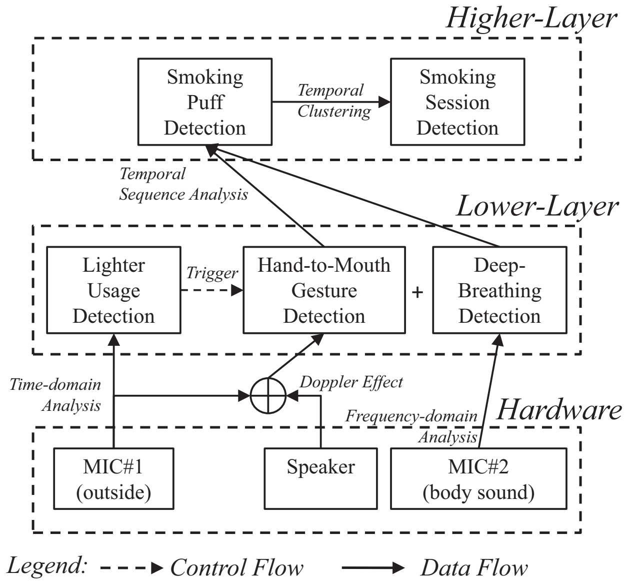


Figure 1: System Overview.

used; 2) the hierarchical framework proposed for data processing fuses the different smoking-related sub-movements which forms unique patterns for smoking and is resistant to confounding events and behaviors.

3. OVERVIEW

We present an overview of the design of the proposed smoking detection system in this section.

3.1 Design Considerations and Principles

As shown above, the three key sub-movements of smoking behavior—lighter usage, deep breathing, and hand-to-mouth gesture have been studied separately in previous work [20, 4, 13, 19, 16] for smoking detection. While each sub-movement may also be present in other confounding behaviors like eating and drinking, the combination of these sub-movements forms a unique pattern of smoking behavior: the smoker often uses the lighter to light the cigarette first, then repetitively feeds the cigarette to the mouth, inhales the smoke, removes the cigarette, and exhales the smoke until the cigarette burns out. As a result, our first design principle is: *model the smoking behavior as a temporal sequential pattern of the subtle sub-movements involving lighter usage, deep breathing, and hand-to-mouth gesture.*

To detect the smoking-related sub-movements, existing work has used different devices for different sub-movements including customized lighter [20], chest/wrist bands [4, 16], etc. However, the extensive use of specialized devices is complex and difficult to wear. Moreover, due to the fact that the user may only wear one wrist band on the dominant hand, existing approach cannot capture the diverse smoking-related gestures which may involve both the dominant and non-dominant hand. As a result, our second design principle is: *design a simple and compact detection device that can accurately detect the smoking-related sub-movements performed in diverse ways.*

Following the above design principles, we introduce the overview of our hierarchical smoking detection framework in the next section.

3.2 System Overview

To achieve simple and reliable smoking behavior detection in daily life, the proposed system adopts a single smart neckband with passive and active acoustic detection abilities to detect the sub-movements in the lower-layer. Smoking puff detection and cigarette counting is then performed in the higher-layer using temporal logic and clustering techniques, respectively. Fig. 1 illustrates an overview of the system's hierarchical architecture.

A smart neckband is designed and implemented combining a smartwatch and a bluetooth microphone. It is equipped with one speaker and two microphones—one facing outside to collect air conducted audio signal (MIC#1); the other, augmented with a stethoscope, facing the throat to collect inner body sound (MIC#2). Detailed hardware design is presented in Sec. 4.1.

After initiation, the hierarchical smoking detection framework works as follows. First, in the lower-layer, MIC#1 continuously collects environmental sound and the system performs low-cost lighter usage detection by time-domain analysis. After detecting a possible lighter usage event, the system triggers the speaker to play a high frequency tone and MIC#2 to record the inner body sound. The speaker and MIC#1 forms the active acoustic detection system for the hand-to-mouth gesture based on the Doppler effect, while MIC#2 detects deep-breathing passively using frequency-domain analysis. In general, the system's lower-layer processes audio data from both microphones to detect the smoking-related sub-movements. Details of the lower-layer system design is introduced in Sec. 4.2.

In the higher-layer, the system processes the temporal sequence of detected sub-movements and detects smoking puffs and smoking sessions. Smoking puff detection is achieved by checking the sequence of sub-movements using a smoking puff model built using Allen's interval algebra [5]. Smoking session detection is then completed by applying temporal clustering over the smoking puffs. Sec. 4.3 presents the details of the higher-layer system.

The system design meets our design principles in the following ways: 1) the smart neckband is simple, compact and easy to wear; 2) the system is fully automated and does not require any user participation before, during, or after smoking; 3) the system well adapts to behavior diversity for the active acoustic detection system can capture the hand-to-mouth gesture regardless the hand used; 4) the hierarchical system is reliable against behavior diversity and confounding events for it fuses different sub-movements that form unique patterns for the smoking behavior.

4. DETAILED SYSTEM DESIGN

We present the details of the hardware and the hierarchical system design in this section.

4.1 Hardware Design and Implementation

The hardware is designed to provide a simple and compact solution to capture essential data for the lower-layer system to detect smoking-related sub-movements including lighter usage, deep-breathing, and hand-to-mouth gesture.

Inspired by existing work on non-speech body sounds [18, 23], deep-breathing detection can be achieved by capturing the breathing sound in the throat by attaching a stethoscope

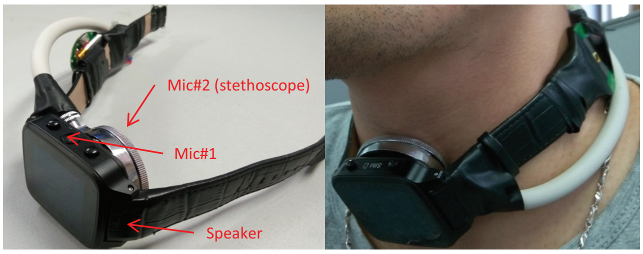


Figure 2: Hardware Design and deployment.

augmented microphone to the throat. The hand-to-mouth gesture performed during smoking involves the hand movement of approaching and departing the mouth to feed the cigarette to the mouth and take it away after inhaled the smoke, respectively. Inspired by existing work on acoustic ranging [17], we design an audio-based active detection system composed of a speaker and a microphone facing outside to detect the hand-to-mouth gesture. Finally, the usage of a lighter creates a unique sound pattern that can be discriminated from other events and noise in the environment, based on which we can perform lighter usage detection.

The above ideas lead to the designment of our smart neckband which provides a pure audio-based sensing solution for smoking-related sub-movements. As shown in Fig. 2, our smart neckband involves two microphones (MIC#1 and MIC#2) and a speaker. MIC#1 faces outside and is responsible of detecting environmental sound, especially for lighter usage. The speaker, combined with MIC#1, forms the active acoustic detection system to capture the Doppler effect caused by the hand approaching and departing the mouth. Finally, MIC#2, augmented with a stethoscope, is designed to capture the inner body sound, especially for deep-breathing detection.

The smart neckband is implemented based on a Cross Country Smartwatch [2] with a dual-core 1.5GHz CPU, 1GB RAM, and the Android 4.2 OS. The built-in speaker and microphone of the smartwatch are capable of playing and recording audio signal up to 22kHz, which we use as the speaker and MIC#1 for our system. MIC#2 is implemented using a stethoscope augmented Bluetooth microphone with a sampling rate of 8kHz. The audio data of MIC#2 is wirelessly streamed to the smartwatch. The smartwatch then processes the data from both microphones and controls the speaker following the design of the hierarchical smoking detection framework introduced in the following sections.

4.2 Sub-movement Detection

This section introduces the design of the lower-layer system which aims at detecting the smoking-related sub-movements from audio data.

4.2.1 Lighter Usage Detection

The first step of the system's lower-layer processing is lighter usage detection, which is achieved by continuously monitoring the ambient sound using the microphone facing outside (MIC#1 in Fig. 2). Because detection is done continuously, it must be light-weight and real-time. Also, because it is a trigger for the remaining parts of the system, it must be optimized for recall (low missing rate). The precision, on the other hand, is less important for even if the system mistakenly regards an irrelevant event as lighter usage (false alarm), it can easily be corrected later for not detecting the subsequent gestures and deep-breathings.

Following the above design trade-off, we propose a very light-weight rule-based lighter usage detection algorithm based on time-domain features. By observing the audio signals of using different lighters (flint or electric ignition lighter), we discover that they all create sudden fluctuations with large amplitudes which can be clearly identified. As a result, we characterize the received audio signal using time-domain features only. For the continuous input audio stream, we first apply a 0.5s non-overlapping sliding window to segment the stream into frames. Time-domain features including the standard deviation (STD) and peak value (Peak) are computed for each frame. Fig. 3 illustrates the instances of lighter usage in the feature space. Fig. 3 also plots instances of similar events such as coughing and speaking. This figure suggests instances of lighter usage form a cluster that can be discriminated from other similar events.

Based on the above observation, we train a lighter usage detector by finding a rectangle in the feature space that covers all lighter usage instances in the training data, as shown in Fig. 3. When a new frame is received, we simply compare its position in the feature space against the rectangle. If it falls into the range of the rectangle, a lighter usage event is detected. Though this simple approach may have a lower precision, it well meets our design trade-off by being very light-weight and fast. The computational complexity for time-domain feature extraction and event detection is linear to the size of the input frame. We also slightly expand (times 1.2) the size of the rectangle to guarantee a high recall.

On detecting the lighter usage event, the system triggers the hand-to-mouth gesture and deep-breathing detection units introduced next.

4.2.2 Hand-to-mouth Gesture Detection

As introduced above, our system detects the hand-to-mouth gesture by detecting the Doppler Effect caused by the hand approaching and departing the mouth when feeding the cigarette to the mouth and taking it away. More specifically, as shown in Fig. 4, the speaker plays a tone with fixed frequency f, the microphone facing outside (MIC#1 in Fig. 2) picks up the audio signal and assume the received frequency is f'. When the hand is close to the smart neckband, part of the emitted audio signal is reflected by the palm. As a result, the received frequency f' is affected by the relative movement between the hand and the device according to the Doppler effect as follows.

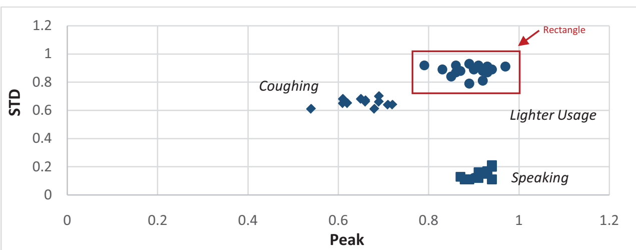


Figure 3: Instances in the feature space.

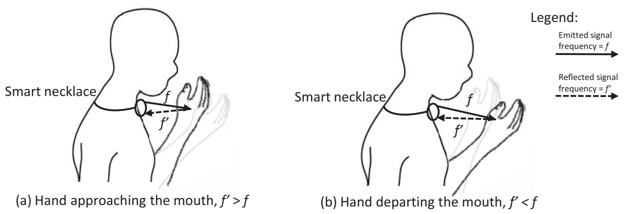


Figure 4: Gesture detection based on Doppler effect.

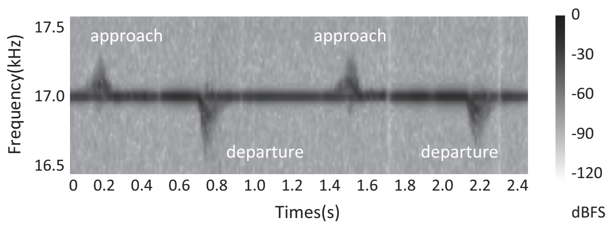


Figure 5: Spectrogram of hand approaching and departing the mouth.

$$f' = (1 + \frac{\Delta v}{c})f \qquad (1)$$

where f' and f are respectively the observed and emitted frequency, c is the velocity of the signal wave traveling in the media, and Δv is the velocity of the palm moving relatively to the speaker, with positive/negative values when the palm is approaching/departing the speaker.

Fig. 5 illustrates the spectrogram of audio signal recorded by MIC#1 with a real-world case of hand approaching and departing the mouth when the user wears the smart neckband. In this case, we set the emission frequency $f = 17kHz$. Clear patterns of frequency shift can be observed. Based on the above analysis and observation, it is straightforward to design the gesture detection algorithm by comparing the signal's energy above and below the emission frequency f. On receiving the raw audio signal, we first apply a non-overlapping sliding window of 0.1s to segment the audio stream into frames. We then apply an 8192 point FFT to get the frequency representation of the frame. The hand is determined to be *approaching*/*departing* the mouth if the signal's energy above f is significantly higher/lower than that below f.

The advantage of this audio-based gesture detection approach lies in that it captures the essential movement of the hand-to-mouth movement during smoking and is not sensitive to the hand used. However, since this approach involves the active emission of an audio signal, collision may take place when multiple devices are present in the same area (e.g., when smokers gather together in the same room). We address the collision problem by two methods: 1) we keep the volume of the emitted signal low so that it can only be detected in a closed range; 2) we implement a simple MAC protocol that each device first sniffers the environment to determine occupied channels before it selects a clear channel above $16kHz$ with $500Hz$ apart from every occupied frequency so that different devices do not interfere with each other and remain inaudible to people.

4.2.3 Deep-breathing Detection

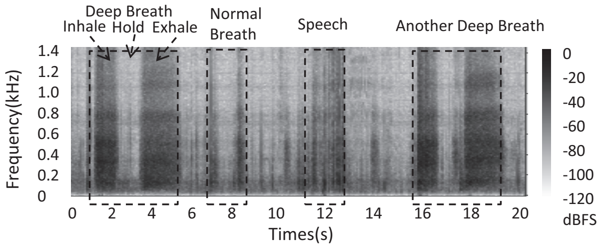


Figure 6: Sound captured by the stethoscope augmented microphone for deep-breathing.

Table 1: Frame Level Features Used

Feature	Description
Spectral Centroid	Center of mass across frequencies
Spectral Flux	Degree of signal change between frames
Spectral Variance	Energy variance across different frequencies
Spectral Skewness	Skewness of spectral distribution
Spectral Slope	The shape of spectra
MFCC	The first 7 Mel Frequency Cepstral Coefficients
Sub-band Energy	Energy of eight different frequency sub-bands
Zero Crossing Rate	Signal diversity in time-domain

Deep-breathing is detected by capturing the non-speech body sound using the stethoscope augmented microphone (MIC#2 in Fig. 2). Deep-breathing is essential for smoking detection because the smoker must inhale and exhale the smoke so nicotine can be delivered to his/her system through the lung. Fig. 6 shows a real case of two deep-breathings apart by a normal breathing and a short speech. It is clear that deep-breathing can be discriminated from other body sounds by analyzing the spectrogram. In this work, we adapt the approach proposed in [18, 23] and perform deep-breathing detection as follows.

Frame Level Features. The system segments the raw audio signal into frames using a 0.1s non-overlapping sliding window. For each frame, we extract twenty-one frame level features as shown in Table 1. The *Sub-band Energy* feature computes the signal's energy in different sub-bands with frequency ranges $(0, f_s/256)$, $(f_s/256, f_s/128)$, $(f_s/128, f_s/64)$, $(f_s/64, f_s/32)$, $(f_s/32, f_s/16)$, $(f_s/16, f_s/8)$, $(f_s/8, f_s/4)$, and $(f_s/4, f_s/2)$, respectively, where f_s is the sampling rate. Detailed explanations are omitted due to page limits, readers can refer to [18, 23] for more information.

Window Level Features. After extracting the frame level features for each frame, we further group the latest frames into a window and compute the window level features. For each of the twenty-one frame level features, we compute its *max*, *min*, *mean*, *variance*, *median*, *number of peaks*, *mean crossing rate*, and *skewness* across the ten frames within the window. As a result, we obtain a feature vector that involves 168 features that represents a window, based on which we perform deep-breathing detection.

Classification. Deep-breathing detection is finally done by using a SVM classifier that classifies the instance represented by window level features into three classes—*deep-breathing inhale*, *deep-breathing exhale*, or *others*.

4.3 Smoking Puff and Session Detection

This section introduces the design of the system's higher-layer that processes the temporal sequence of the sub-movements

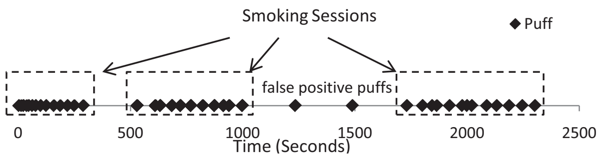


Figure 7: Temporal clustering of smoking puffs

for smoking puff and session detection.

4.3.1 Smoking Puff Detection

Triggered by the lighter usage event, the smoking puff detection unit detects each smoking puff by analyzing the temporal sequence of smoking-related sub-movements—*deep breathing (inhale/exhale)* and *hand-to-mouth gesture (approach/depart)*. After closely observing and interviewing the smokers, we find a smoking puff typically starts with the hand approaching the mouth to feed the cigarette. After the smoker has sucked the smoke into the mouth, the cigarette is removed from the lips by a hand departure gesture while the smoker takes a deep-breath to inhale the smoke into the lung. A smoking puff ends with the smoker exhale the smoke out. Based on the above observation, we model each smoking puff using the above sub-movements by Allen's interval algebra[5]. The model is built by a series of rule which we informally present as follows: 1) *hand approaching* **starts** *puff*; 2) *hand departure* **during** *puff*; 3) *inhale* **during** *puff*; 4) *exhale* **finishes** *puff*.

Smoking puff detection is done by constantly checking the input sub-movement sequence against the above rules. In order to tolerant possible errors in sub-movement detection, a puff is successfully detected as long as three out of the above four rules are met. Also, the duration of one puff is empirically restricted to be within $[1.2, 5.6]$ seconds according to the statistical results from our benchmark data set involving twenty cigarette sessions collected from three smokers under various conditions.

4.3.2 Smoking Session Detection

Based on the smoking puff detection results, we discover cigarette sessions during a period of time by applying temporal clustering [15] over the puffs. As shown in Fig. 7, given the series of smoking puffs ordered by time, we cluster the puffs by their time intervals and find the smoking sessions by the clustering results. We empirically set the threshold of time interval to 60s and each smoking session to involve at least 9 puffs according to the same benchmark data set introduced above.

5. EXPERIMENTS

We present the experiment results to evaluate the proposed system in this section.

5.1 Data Collection and Methodology

Data collection lasts for 7 weeks and we collected a total number of 143 cigarette sessions performed by 16 different subjects. The data collection process is divided into two stages. In the first stage, we ask the smoker to wear our smart neckband when smoking. Each record contains data of one smoking session with possible concurrent activities in different environments. The concurrent activities mainly include *walking*, *drinking*, *eating*, *coughing*, and *laughing*. Different environments are classified based on noise level which include *quiet indoor*, *quiet outdoor*, *noisy indoor*, and *noisy outdoor*. A total number of 14 subjects are involved in this stage and 107 smoking sessions are recorded which last 323.3 minutes. We denote this data set as **DATASET#1**.

In the second stage, we recruit 2 smokers to wear our smart neckband during their normal daily life and log their activities. This record lasts for 1486.5 minutes and contains 36 smoking sessions with other activities performed alone or during smoking. Besides the concurrent activities listed above, we also discover four confounding activities to have similar gesture or audio pattern to smoking including *eating*, *drinking*, *coughing* and *waving hand*. We denote this data set as **DATASET#2**.

Evaluation is done on these two data sets as follows. We evaluate the system's overall performance by combining the two data sets. The impact of environmental noise on the system's performance is evaluated using DATASET#1. We study the influence of concurrent and confounding activities using DATASET#1 and DATASET#2, respectively. We also compare the proposed approach against approaches that use one sensing modality (i.e., gesture or deep-breathing only) to demonstrate the advantage of our system.

The metrics used for performance evaluation include *recall*, *precision*, and F_1 *score* which are defined as follows.

$$recall = \frac{\texttt{True Positive}}{\texttt{True Positive + False Negative}}$$

$$precision = \frac{\texttt{True Positive}}{\texttt{True Positive + False Positive}}$$

$$F_1 score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

The F_1 *score*, defined as the harmonic mean of precision and recall, is used as a unified metric to evaluate the system's performance. The smoking puff and session detection in the system's higher-layer do not involving model training and is evaluated directly on the data set. The performance for sub-movement detection in the system's lower-layer is obtained by ten-fold-cross-validation. Detailed evaluation results are presented in the following sections.

5.2 Overall Performance

This section presents the evaluation result of the system's overall performance.

First, we evaluate the performance of smoking puff detection on the combined data set of DATASET#1 and #2. Smoking puff detection is essential to our system based on which smoking session detection is done by temporal clustering the puffs. Fig. 8 illustrates the *recall*, *precision*

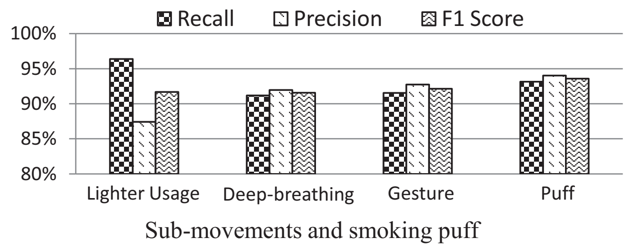
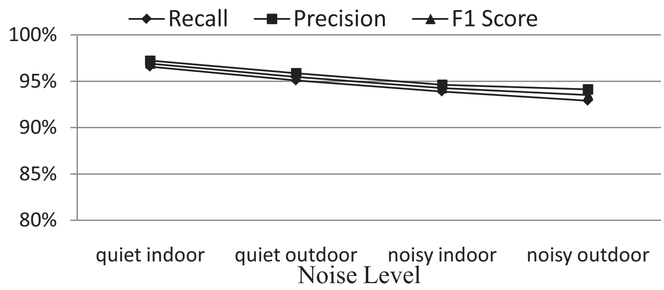
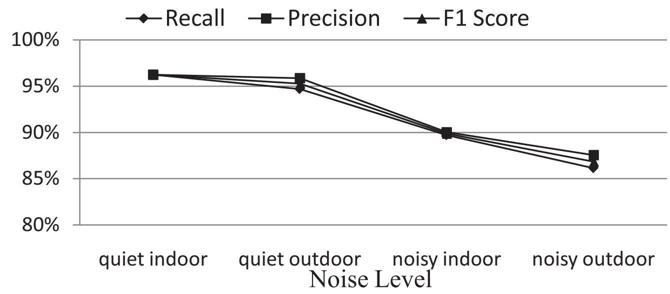


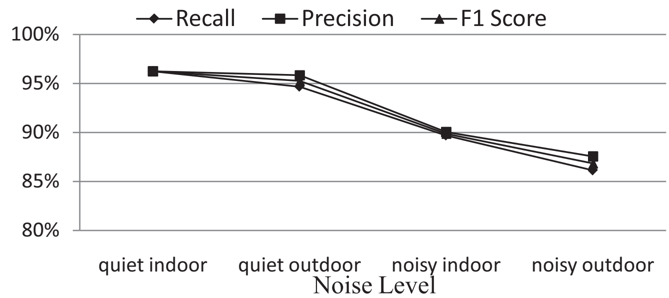
Figure 8: Performance of smoking puff detection.



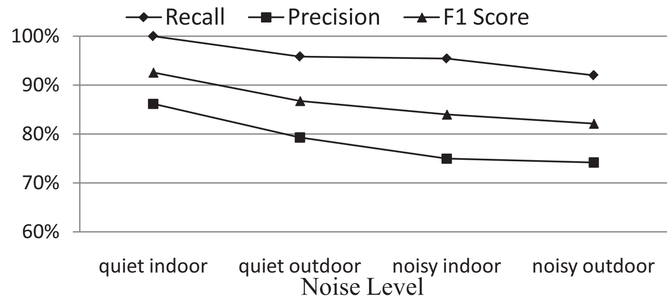
(a) Smoking puff detection



(b) Hand-to-mouth gesture detection



(c) Deep-breathing detection



(d) Lighter usage detection

Figure 9: Impact of environmental noise level on smoking puff and sub-movement detection performance.

Table 2: Performance of smoking session detection.

	Recall	*Precision*	*F1 Score*
Puff	86.78%%	88.48%	87.62%
Session	91.67%	94.29%	92.96%

and F_1 *score* of detection results for smoking puffs and the component sub-movements. This result suggests our system can accurately detect smoking puffs with *recall*=93.15%, *precision*=94.03%, and F_1 *score*=93.59%. The performance of smoking puff detection is better than deep-breathing and gesture detection alone because we fuse these two modalities to detect puffs, demonstrating the advantage of our system design. For lighter usage detection, we achieve a high *recall* of 96.39%, suggesting the light-weight detection algorithm well achieves our design goal. Overall speaking, the system achieves high accuracy detection for smoking puffs and sub-movements with F_1 *score* all above 91%.

Second, we evaluate the performance of smoking session detection based on the smoking puff detection results. We perform the study on DATASET#2 because DATASET#1 is already known to have one session in each record. Table 2 shows the results. This result suggests based on an accurate detection of smoking puffs, our system can achieve high accuracy detection of smoking sessions in smokers' daily lives. This result also suggests the performance of smoking session detection is better than the smoking puff detection, this is because the temporal clustering approach adopted is not sensitive to occasional mistakes involved in the puff detection results, as shown in Fig. 7. Due to the fact that smoking session detection heavily relies on smoking puff detection in our system, for the following experiments, we mainly focus on evaluating the performance of smoking puff detection.

5.3 Impact of Environmental Noises

In this experiment, we evaluate the system's performance in different environments classified using the noise level. There are four types of environments with an increasing noise level namely the *quiet indoor*, *quiet outdoor*, *noisy indoor*, and *noisy outdoor*. DATASET#1 is used for this experiment for it contains smoking sessions taken place in known environments. Information about the four environments is as follows.

The *quiet indoor* environment is a corridor in the lab building with occasional noise of people walking and door opening and closing. The *quiet outdoor* environment is located in a remote stadium on campus with a few students jogging on the track and the noise from an avenue outside the campus. We choose a discussion room as the *noisy indoor* environment where people discuss, typing on the laptops computers with the air conditioning on. Finally, the *noisy outdoor* environment is selected to be at a bus stop by the main street with a lot of cars and buses pass and stop and people talk and laugh while waiting for the buses. From the above description, it is clear that there is an increasing noise level among the above four environments. The 107 smoking sessions involved in DATASET#1 are distributed roughly equally in different environments to draw a fair comparison.

Fig. 9 summarizes the performance of smoking puff and sub-movement detection in different environments. First of all, the system's performance generally decreases with an increasing level of noise. For smoking puff detection, the F_1 *score* drops from 96.91% in the *quiet indoor* environment to 93.51% in the *noisy outdoor* environment as shown in Fig. 9(a). Similar performance drop can also be observed for different sub-movements as shown in Fig. 9(b)-(d). For gesture detection, the F_1 *score* drops from 96.23% to 86.83%.

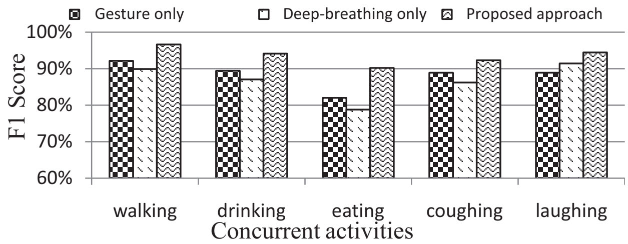


Figure 10: Smoking puff detection performance with concurrent activities.

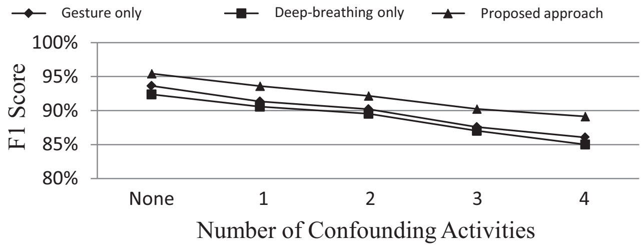


Figure 11: Smoking puff detection performance with confounding activities.

And for deep-breathing detection, the F_1 *score* drops from 96.05% to 89.29%. For lighter usage detection, the most important metric is the *recall* for being the trigger of the system's remaining parts as discussed above. The *recall* of lighter usage detection drops from 100% to 92% with noise level increasing from *quiet indoor* to *noisy outdoor*. This result suggests that, though affected by noise, our system achieves high detection accuracy for both the smoking puffs and the sub-movements even in a very noisy environment.

Next, from Fig. 9(b) and (c), it is clear that the *recall* and *precision* are always close for gesture and deep-breathing detection results in different environments. This result suggests false positive (false alarm) and false negative (miss detection) increase in similar scale with the increasing level of noise for these two sub-movements. However, by comparing the results in Fig. 9(a) and Fig. 9(b)-(c), it is clear that the performance drop in smoking puff detection is much less significant than those for gesture and deep-breathing detection. This is because we detect smoking puffs by matching the majority of temporal logic rules over the sub-movement sequence as introduced in Sec. 4.3. As a result, the increase of false alarm and miss detection of gesture and deep-breathing can affect the final puff detection result only if they happen simultaneously, which is much less likely comparing to puff detection based on a single modality.

In summary, our system achieves reliable performance in noisy environments, and by fusing different sensing modalities, achieves better performance for smoking puff detection than single sub-movements. Smoking session detection result is not available in this experiment because each record is known to have only one session in the data set used. However, according to the high performance of smoking puff detection and the result presented in Table 2, it is reasonable to expect that we can achieve high smoking session detection accuracy in different environments.

5.4 Concurrent Activities

Smokers may carry out other activities when smoking, e.g., waling, drinking, eating, etc. Such concurrent activities can be observed in both DATASET#1 and #2 for short- and long-term data collection, respectively. From these data sets, we discover five activities that are frequently performed as concurrent activities to smoking, including walking, drinking, eating, coughing, and laughing. We resample the data sets to form a new data set that involves 216 smoking sessions with concurrent activities.

Fig. 10 shows the performance (F_1 *score* as metric) of smoking puff detection with concurrent activities and compares the result of using the proposed fusing approach against using only one modality (i.e., only deep-breathing or gesture). As shown in the figure, our approach outperforms other approaches that uses only one modality for smoking puff detection. Among all the concurrent activities, *eating* is shown to have the most significant influence on the system's performance. It is because *eating* has the similar hand-to-mouth gesture for bringing the food to the mouth and deep-breathing pattern after swallowing the food. When the smoker smokes while eating, the proposed system has the lowest performance with F_1 *score*=90.2%. For smoking puff detection only relies on gesture or deep-breathing, the F_1 *score* is 82% and 78.79%, respectively.

In summary, with the presence of concurrent activities, the proposed system achieves a smoking puff detection accuracy of above 90% in F_1 *score*. Compared to approaches that use only one modality for puff detection, the proposed system shows an overall improvement of 6.3% in F_1 *score*, and the largest improvement is 11% in the *eating* case.

5.5 Confounding Activities

Confounding activities are activities with similar gesture and breathing patterns that can easily be confused with smoking. Unlike concurrent activities, confounding activities do not necessarily happen simultaneously with smoking. With a high performance lighter usage detection as shown in Fig. 8 and Fig. 9, confounding activities can largely be eliminated because lighter usage is unique for smoking than other activities. However, since the lighter usage detection unit is optimized for *recall* and light-weight, the system can be occasionally triggered by a false positive detection. In this section, we evaluate the system's performance with the presence of confounding activities by assuming the system is triggered by false positive lighter usage detection results.

We use DATASET#2 that involves data of 2 smokers over long periods to evaluate the system's performance with confounding activities. After closely observing the data, we list four activities that frequently occur and similar to smoking, i.e., *eating*, *drinking*, *coughing* and *waving hand*. In this data set, there are a total number of 36 smoking sessions involving 531 puffs and 152 instances of the above confounding activities. To study the impact of confounding activities on the system's performance, we resample the data set to have it contains *no*, only *one* confounding activity, to all *four* confounding activities.

Fig. 11 illustrates the system's performance on smoking puff detection with the presence of different number of confounding activities. As shown by the figure, the system's performance smoothly decreases with the increasing number of confounding activities. When there is no confounding activity, the system achieves the highest performance of F_1 *score*=95.45%. With all four confounding activities present, the F_1 *score* drops to 89.15%. Despite the performance drop, the system still maintains high detection accuracy with the presence of confounding activities.

Moreover, from Fig. 11, it is clear that the proposed system outperforms approaches that uses one modality for smoking puff detection in all cases by at least 3%.

In summary, the multi-modal smoking puff detection approach proposed in this paper can still accurately discriminate smoking puffs from other confounding activities. Combining the overall result presented in Table 2, we conclude that the system can accurately detect smoking puffs and sessions during smokers' daily lives with the presence of concurrent and confounding activities.

5.6 Computational Overhead and Power Consumption

In this section, we evaluate the overhead of our system implemented on Android watch and a Bluetooth module. We analyze data using offline methods, the Android watch take 2.9J per minute that is around 48 mW. The Bluetooth module we use the modified Rocketfish Bluetooth headset, the headset can last over 12 hours when it is recording or playing music. The Android smart phone is added for solve the Bluetooth protocol's problem, in the future, our system can drop it, so in this paper, we do not care about the overhead of this phone, besides it cost less than Android watch. Overall, we can find because we choose lightweight design, our system has low battery usage, despite power consumption may be different on different Android watches, we argue that we present a novel and useful wearable device to detect smoking in daily life, with the improvement of hardware design, we trust the neckband cost lower energy consumption.

6. CONCLUSION

This paper presents the design and implementation of a simple and compact smart neckband for smoking behavior detection. The smart neckband include both the passive and active acoustic detection system that captures the smoking-related sub-movements, i.e., lighter usage, hand-to-mouth gesture, and deep-breathing, as the key indicators of smoking behavior. The hierarchical processing framework involve two layers. The lower-layer detects lighter usage as the trigger for the system's remaining parts by a light-weight time-domain analysis algorithm. The hand-to-mouth gesture detection and deep-breathing detection are also done by the lower-layer by analyzing the Doppler effect of audio signal reflected by the palm and pattern recognition techniques, respectively. Given the sub-movement sequence detected by the lower-layer, the higher layer first perform smoking puff detection by checking the sequence against the puff model described by a series of temporal rules. Smoking session is then detected by applying temporal clustering over the smoking puff sequence. Experiment results suggest the proposed system can accurately detect smoking puffs and sessions in real life scenarios. The system has also shown to be reliable even with a high level of environmental noise and the presence of multiple concurrent and confounding activities.

For our future work, we plan to first further reduce the size and cost of our smart neckband by only preserving the hardware necessary for data transmission and high quality audio play and record. The data processing will be migrated to a smartphone which is commonly available. The current Bluetooth-based devices (e.g., headsets) are insufficient in performing the required duplex high quality audio data transmission. As a result, a WiFi-based device may be an option. Next, we plan to design and develop a smartphone application that will 1) present the smoking detection results in a user friendly way; and 2) motivate the smokers to reduce even quit smoking. We also plan to use sequential data mining technologies to improve the performance of our system.

Moreover, as a general platform equipped with both passive and active acoustic devices, the proposed smart neckband can be used for many other potential applications besides smoking behavior analysis. First, our platform can be used to detect various non-speech body sounds [18, 23]. Based on our smart neckband, it is possible to build applications such as eating and drinking behavior analysis for eating disorder or calory intake studies. By fusing the breathing data with other inertial sensor readings on the smart neckband, we can build applications including sport aid and lung function analysis. Second, by exploring the potential of the active acoustic system, this platform can be used for applications such as capturing face-to-face interactions [11], estimating the distance between different subjects [17], and recognizing gestures performed in front of the user, etc. We plan to build other possible applications using our platform in the future.

Acknowledgments

We thank all the volunteers participated in our study and the anonymous reviewers that help us to improve this work.

This work is supported by National 863 Program No. 2015AA01A203; NFSC under Grants 61502225, 61373011, 91318301; and the Collaborative Innovation Center of Novel Software Technology and Industrialization.

7. REFERENCES

[1] Cress. http://www.borgwaldt.de/cms/borgwaldt-kc/produkte/rauchmaschinen/geraete-zur-rauchtopographie/cress-pocket.html.

[2] Cross country smartwatch. http://www.ebay.com/itm/171529487253.

[3] Smokerlyzer. http://www.bedfont.com/smokerlyzer.

[4] A. A. Ali, S. M. Hossain, K. Hovsepian, and M. M. Rahman. mpuff: Automated detection of cigarette smoking puffs from respiration measurements. In *IPSN*, pages 269–280, 2012.

[5] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of The ACM*, 26(11):832–843, 1983.

[6] J. Bryant, B. Bonevski, C. Paul, and C. Lecathelinais. Assessing smoking status in disadvantaged populations: is computer administered self report an accurate and acceptable measure? *Bmc Medical Research Methodology*, 11(1):1–8, 2010.

[7] N. C. for Chronic Disease Prevention, H. P. U. O. on Smoking, and Health. *The Health Consequences of Smoking–50 Years of Progress: A Report of the Surgeon General*. United States. Public Health Service. Office of the Surgeon General., 2014.

[8] R. M. Glass. Blue mood, blackened lungs. depression and smoking. *Jama the Journal of the American Medical Association*, 264(12):1583–4, 1990.

[9] D. Hammond, G. T. Fong, K. M. Cummings, and A. Hyland. Smoking topography, brand switching, and nicotine delivery: results from an in vivo study. *Cancer Epidemiology Biomarkers & Prevention*, 14(6):1370–5, 2005.

[10] R. I. Herning, R. T. Jones, J. Bachman, and A. H. Mines. Puff volume increases when lownicotine cigarettes are smoked. *British Medical Journal*, 283(6285):187–189, 1981.

[11] W. Huang, Y. S. Kuo, P. Pannuto, and P. Dutta. Opo: a wearable sensor for capturing high-fidelity face-to-face interactions. In *ACM Conference on Embedded Network Sensor Systems*, pages 61–75, 2014.

[12] S. Kolonen, J. Tuomisto, P. Puustinen, and M. M. Airaksinen. Puffing behavior during the smoking of a single cigarette in a naturalistic environment. *Pharmacology Biochemistry & Behavior*, 41(4):701–706, 1992.

[13] P. Lopez-Meyer, S. Tiffany, and E. Sazonov. Identification of cigarette smoke inhalations from wearable sensor data using a support vector machine classifier. In *IEEE Engineering in Medicine & Biology Society Conference*, pages 4050–3, 2012.

[14] N. Moghaddam and E. Ferguson. Smoking, mood regulation, and personality: An event sampling exploration of potential models and moderation. *Journal of Personality*, 75(3):451, 2007.

[15] N. Molinari, C. Bonaldi, and J. P. Daures. Multiple temporal cluster detection. *Biometrics*, 57(2):577, 2001.

[16] A. Parate, M. C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis. Risq: recognizing smoking gestures with inertial sensors on a wristband. In *MobiSys*, pages 149–161, 2014.

[17] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *International Conference on Embedded Networked Sensor Systems*, pages 1–14, 2007.

[18] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury. Bodybeat: A mobile system for sensing non-speech body sounds. In *MobiSys*, pages 2–13, 2014.

[19] E. Sazonov, K. Metcalfe, P. Lopez-Meyer, and S. Tiffany. Rf hand gesture sensor for monitoring of cigarette smoking. In *Sensing Technology (ICST), 2011 Fifth International Conference on*, pages 426–430, 2011.

[20] P. M. Scholl, N. Kĺźcĺźkyildiz, and K. V. Laerhoven. When do you light a fire?: capturing tobacco use with situated wearable sensors. In *ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pages 1309–1310, 2013.

[21] D. Shipton, D. M. Tappin, T. Vadiveloo, J. A. Crossley, D. A. Aitken, and J. Chalmers. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *Bmj*, 339(oct29 1):1241–1242, 2009.

[22] J. C. Veilleux, J. D. Kassel, A. J. Heinz, A. R. Braun, M. C. Wardle, J. E. Greenstein, D. P. Evatt, and M. Conrad. Predictors and sequelae of smoking topography over the course of a single cigarette in adolescent light smokers. *Journal of Adolescent Health*, 48(2):176–181, 2011.

[23] K. Yatani and K. N. Truong. Bodyscope: a wearable acoustic sensor for activity recognition. *UbiComp*, pages 341–350, 2012.

[24] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu. Smokey: Ubiquitous smoking detection with commercial wifi infrastructures. In *IEEE INFOCOM 2016 - IEEE Conference on Computer Communications*, 2016.