Tauhidur Rahman *Cornell University* **Alexander T. Adams** *University of North Carolina at Charlotte*
Mi Zhang *Cornell University* **Erin Cherry** *University of Rochester* **Tanzeem Choudhury** *Cornell University*

BODYBEAT: Eavesdropping on our Body Using a Wearable Microphone

From munching on a piece of toast and swallowing a sip of coffee to deep breathing after a few laps of running, our body continually makes a wide range of non-speech body sounds, which can be indicative of our dietary behaviour, respiratory physiology, and affect. A wearable system that can continuously capture and recognize different types of body sound with high fidelity can also be used for behavioural tracking and disease diagnosis. BodyBeat is such a mobile sensing system that can detect a diverse range of non-speech body sounds in real-life scenarios. The BodyBeat mobile sensing system consists of a custom-built piezoelectric microphone and a distributed computational framework that utilizes an ARM microcontroller and an Android smartphone. The custom-built microphone is designed to capture subtle body vibrations directly from the body surface without being disturbed by external sounds. The ARM embedded system and the Android smartphone processes the acoustic signal from the microphone and identifies non-speech body sounds.

Speech is not the only sound generated by human. Non-speech body sounds such as sounds of food intake, breath, laughter, yawn, and cough contain invaluable information about people's health and wellbeing. With regard to food intake, body sounds enable us to discriminate characteristics of food and drinks [1, 2]. Longer term tracking of eating sounds could be very useful in dietary monitoring applications. Breathing sounds, generated by the friction caused by the airflow from our lungs through the vocal organs (e.g., trachea, larynx, etc.) to the mouth or nasal cavity [3], are highly indicative of the conditions of our lungs. Sounds of laughter and yawns are good indicators of people's affect states such as happiness and fatigue. Therefore, automatically tracking these non-speech body sounds can help in early detection of negative health indicators by performing regular dietary monitoring, pulmonary function testing, and affect sensing.

We have designed, implemented, and evaluated a mobile sensing system called BodyBeat, which could continuously keep tracking of a diverse set of non-speech body sounds. BodyBeat consists of a custom-made piezoelectric sensor-based microphone, an ARM microcontroller, and an Android smartphone application. The piezoelectric microphone captures body sounds and in the meantime dampens external sounds such as human speech sound and ambient noises. A classification algorithm is developed based on a set of acoustic features to recognize different types of body sounds and is implemented on the ARM micro-controller and the Android smartphone application. We have extensively evaluated the performance of BodyBeat. We have also evaluated the performance of the body sound classification algorithm, and profiled the system performance in terms of CPU and memory usage and power consumption.

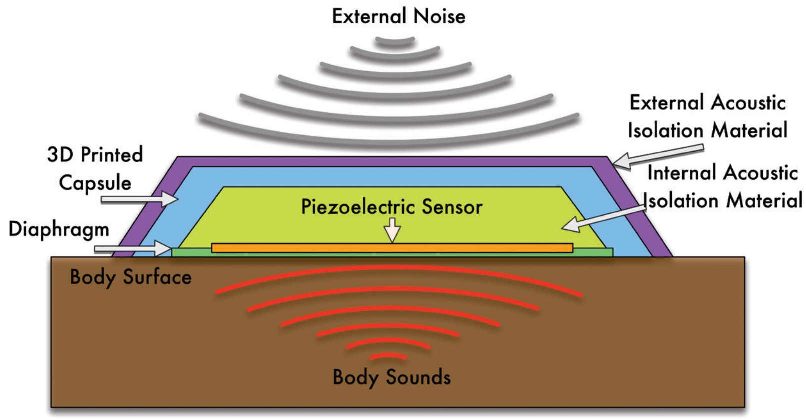

FIGURE 1. Proposed BodyBeat Microphone Architecture.

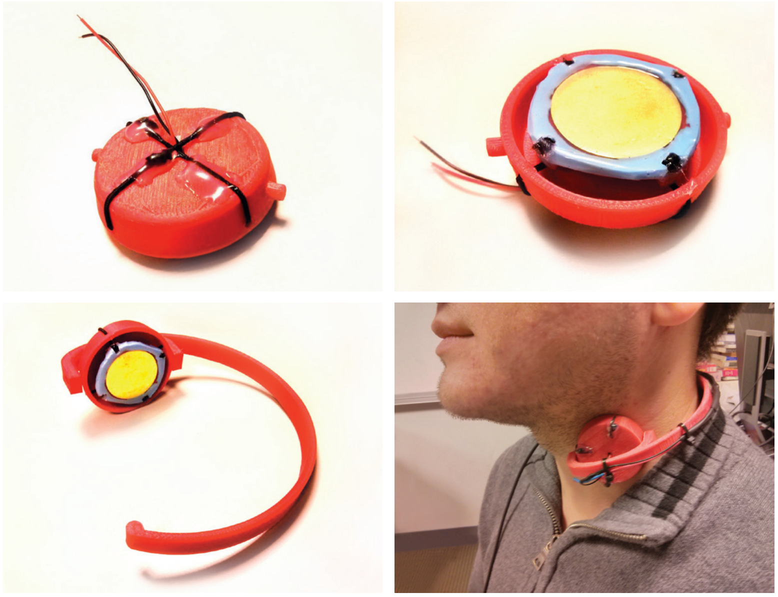

FIGURE 2. The top view (top left) and bottom view (top right) of the microphone with suspension capsule. It also shows the neckpiece design (bottom left) and how it fits on a person's neck (bottom right).

MICROPHONE DESIGN:

Non-Speech body sounds are generated by complex physiological processes and must traverse through a series of tissues, fluids, and bone before getting to air, which cause these sounds to lose a lot of energy. These subtle sounds are also mostly found in the lower end of the audio spectrum making the microphone on smartphones not ideal for capturing these sounds. When designing a microphone optimized for capturing non-speech body audio, the following should be considered:

- Maximizing the transfer of energy between different mediums (body to capsule)
- High Sensitivity in the frequency domain of the desired non-speech body audio
- External and ambient noise rejection
- Compensation for friction noise generated by users' movements

The BodyBeat microphone is a customized contact microphone system that is worn around the user's neck, giving access to a wide array of non-speech body sounds that lie between 20Hz and 1300Hz. In addition to optimizations to capture these sounds, the BodyBeat microphone is also designed in a way that rejects external and ambient noises in addition to being robust against friction/movement noise.

As can be seen in Figure 1, the microphone consists of a brass piezoelectric disk, which is covered in soft silicone, housed in a 3D printed (ABS) capsule. This capsule is filled with silicone and also covered in more dense silicone, both of which aid in acoustic isolation, and it's suspended inside a larger capsule via elastic suspension cables. The suspension cables, which connect the internal capsule to the larger external capsule, help minimize friction noise generated by movement, allowing silicone coating on the piezo-element to stick to the surface of users' skin while the neck piece and outer capsule can move around it. The layer of silicone that coats the piezo-element has similar characteristics to human skin, which helps with energy loss due to having similar impedance. Figure 2 shows the neckpiece that is built on top of the basic BodyBeat microphone.

CLASSIFIER DESIGN:

To train and evaluate a supervised classifier for a wide range of body sounds, 14 participants were recruited. We collected different kinds of body sounds capturing

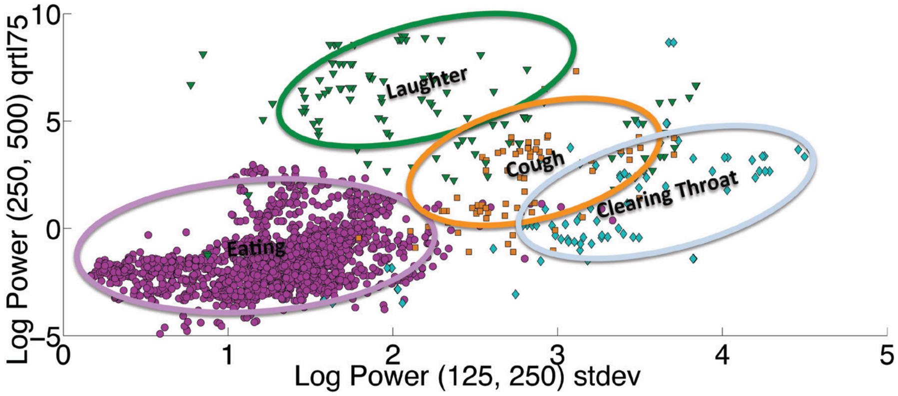

FIGURE 3. Four types of body sounds in a feature space formed by two top selected features by feature selection algorithm.

eating and drinking events (i.e. eating chips, bread, apple, and drinking water), cardiopulmonary health (i.e. cough, clearing throat, deep breathing) and affect (i.e. laughter). We also collected some non-body sounds (i.e. silence and speech). Every participant contributed all these different kinds of body sounds, which sums up to approximately 15 minutes of body sound recordings with a sampling rate of 8 kHz and 16 bit resolution. All these recordings are conducted in a laboratory environment while the microphone of our BodyBeat neckpiece prototype is placed near the vocal chord.

A two-step feature extraction procedure is employed to characterize body sounds. Firstly, a set of frame-level acoustic features is extracted that captures interesting spectral and energy pattern in the low-level. Our initial frame-level feature set contains 8 sub-band power features, RMS energy, zero crossing rate (ZCR), 9 spectral features, 12 Mel Frequency Cepstral Coefficients (MFCCs). In the second step, based on those extracted frame-level features, we grouped frames into windows and extract features at the window-level. We applied a set of statistical functions across all the frame-level features within each window to capture the averages, extremes, rate of change, and shape of the frame-level features within each window. Upon extracting an initial set of 512 window-level features, we adopt the Correlation-based Feature Selection (CFS) algorithm to select a subset of features [4]. We selected a set of 30 window level features with the CFS algorithm. Figure 3 shows a scatter plot of 4 different types of body sounds in a feature space formed by two top selected features.

As our supervised body sound classifier, we use a Linear Discriminant Classifier (LDC). Two different cross-validation experiments: a Leave-One-Person-Out (LOPO) and a Leave-One-Sample-Out (LOSO) cross-validation experiment, is conducted to evaluate the performance of our proposed classifier. Table 1 shows the average performance (in terms of Recall, Precision and) of different classifiers with different feature sets. We explored the performance of this LDC with different combination of energy, spectral features, and MFCC for selecting the top window-level features. As can be seen in Table 1 the performance reaches to its peak (recall of 72.5% and precision of 63.4%), when we used all the frame level features. However, with only energy and spectral features as frame-level features, the LDC yields a nice performance with a recall of 71.2% and precision of 61.5% from the LOPO experiment. Considering the computational complexity of MFCC feature, we decided to use just energy and spectral features as frame-level features with LDC as classifier for the rest of our analysis and system implementation. Lastly, our classifier analysis also shows that this LDC with our optimized feature set outperforms the classification algorithm proposed by a recent study [2]. This performance also illustrates the potential of a piezo-electric sensor-based wearable microphone for continuous monitoring of different body sounds. The difference between LOSO and LOPO also shows that personalization of such classifiers could significantly improve the performance of the classifiers.

TABLE 1: Performance of the BodyBeat classifier in terms of Recall (R), Precision (P) and F-measure (F).

Frame-level Features	LOPO			LOSO		
	R	P	F	R	P	F
Energy & Spectral	71.2	61.5	66.5	88.1	81.9	86.5
MFCC	66.3	52.8	57.8	75.0	71.5	73.2
Energy & Spectral & MFCC	72.5	63.4	67.6	90.3	82.3	86.6
BodyScope [2]	57.6	55.5	56.5	76.6	71.5	73.8

SYSTEM IMPLEMENTATION:

The BodyBeat non-speech body sound sensing mobile system (Figure 4) is implemented using an embedded system unit and an Android application unit. The custom-made microphone of BodyBeat system is directly attached to the embedded system. The embedded system unit utilizes an ARM microcontroller unit, an audio codec and a Bluetooth module to implement capture, preprocessing and frame admission control of the raw acoustic data from the microphone. At the center of the embedded system unit, we used a ARM microcontroller [5]. The board consists of a 72MHz ARM cortex M3 chip. The ARM microcontroller connects to an audio codec via SPI [6]. The ARM microcontroller also does a frame admission control to filter out audio frames that do not contain any body sounds. After getting the FFT of the

Hanning windowed audio frame of 1024 samples, we extracted a few important sub-band power and zero crossing rate features to detect the presence of speech and silence. Upon processing the raw audio data, the ARM unit sends data to the Android phone using a class 2 Bluetooth modem (RN-42 chip) with an SPP profile. As the Android application unit receives data packets from the embedded system unit, two dimensional circular buffer containing 24 data frames gets filled for the 2 stage feature extraction and inference. The entire feature extraction and classification algorithm is implemented in the native layer considering the speed requirements for real-time passive body sound sensing.

Lastly, we benchmarked both the Android and the embedded unit. When the BodyBeat microphone captures either silence or speech, the Android application unit consumes less than 12% of the CPU and 45 MB of memory, because of embedded system's frame admission control. During the presence of body sounds, the CPU and memory usage increases and reaches up to 22% and 47 MB. The embedded system unit consumes about 333.3 mW while the raw audio data contains valuable body sounds and the frame admission control allows the data to be transferred to the Android system unit. On the other hand, when frame admission control detects either silence or speech, the embedded system unit's power consumption decreases to 289.971mW. The average power consumption by the Android application unit is about 374.49 mW and could easily last a fully charged Nexus 5 about 10 hours.

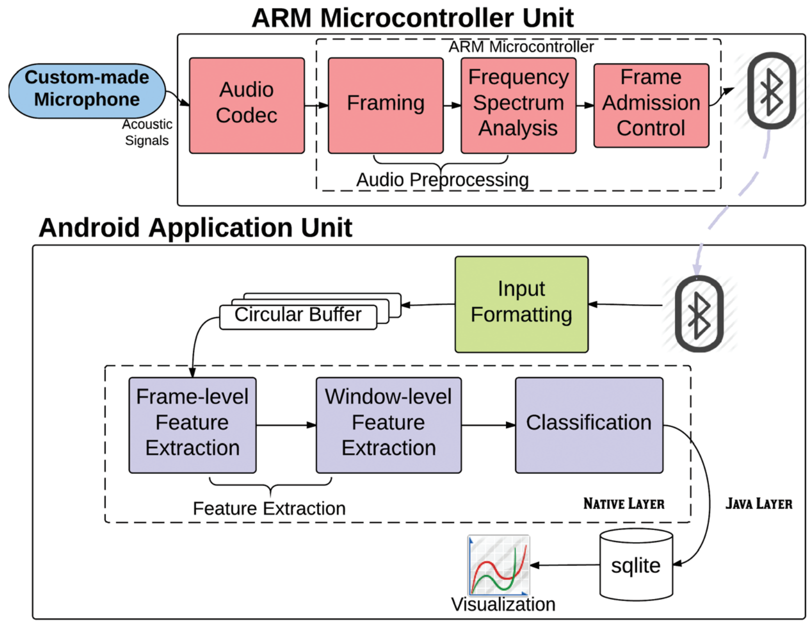

FIGURE 4. System architecture with both the ARM microcontroller unit and the Android application unit.

CONCLUSION AND FUTURE APPLICATIONS

Wearable mobile systems are playing a key role to revolutionize health of the masses. Different wearable systems are increasingly being used for early and cheap diagnosis of diseases and for tracking different health variables. Inspired by the same philosophy, we proposed BodyBeat in this work, which is a mobile system for continuous monitoring of different kinds of body sounds. We discuss the microphone design, classifier design and system implementation in an Android smartphone and an ARM microcontroller. For a detailed evaluation of our proposed microphone, signal processing, machine learning algorithm and mobile system, please refer to the original paper [9].

Continuous and passive detection of body sounds in a wearable form factor can provide us with new opportunities to build novel mobile health applications. One of the applications of BodyBeat could be in automatic food journaling. Tracking daily eating behavior is a very challenging problem that managed to attract the attention of a lot of researchers. Some studies aim to solve the problem with a camera and computer vision algorithms while some studies resort to online crowdsourcing platform (e.g. PlateMate[7]). As BodyBeat can recognize eating and drinking sounds, it has the potential to be used in automatic or semi-automatic food journaling applications. Upon detecting the moments when user is eating, the system can use Microsoft SenseCam or Google Glass to automatically take a picture of her food for further analysis. The BodyBeat also enables us to detect coughing, deep or heavy breathing, which can be indicative of many pulmonary diseases. While a few previous studies have illustrated success detecting these body sounds indicative of illness [2, 8], BodyBeat mobile system can be used in an application which will detect the onset, frequency, and the location of coughing, heavy breathing or any other kind of pulmonary sounds. ■

REFERENCES

[1] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster. Analysis of chewing sounds for dietary monitoring. In UbiComp 2005, 56-72.
[2] K. Yatani and K. N. Truong. BodyScope: a wearable acoustic sensor for activity recognition. In UbiComp 2012, 341-350.
[3] S. Reichert, R. Gass, C. Brandt, and E. Andrès. Analysis of Respiratory Sounds: State of the Art. In Clinical Medicine 2008. Circulatory, Respiratory and Pulmonary Medicine, 2, 45–58.
[4] Hall, M. A. Correlation-based Feature Selection for Machine Learning. PhD Thesis April 1999.
[5] http://leaflabs.com.
[6] http://www.openmusiclabs.com/projects/codec-shield/.
[7] J. Noronha, E. Hysen, H. Zhang, and K. Gajos. Platemate:crowdsourcing nutritional analysis from food photographs.In UIST 2011.
[8] E. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. Patel. Accurate and privacy preserving cough sensing using a low-cost microphone. In UbiComp 2011, 375–384.
[9] 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 2014. 2-13.