2017 IEEE 23rd International Conference on Parallel and Distributed Systems

BARTON: Low Power Tongue Movement Sensing with In-ear Barometers

Balz Maag*, Zimu Zhou*, Olga Saukh† and Lothar Thiele*
*Computer Engineering and Networks Laboratory, ETH Zurich
†Graz University of Technology / Complexity Science Hub Vienna
*{balz.maag, zimu.zhou, thiele}@tik.ee.ethz.ch
†saukh@tugraz.at

***Abstract*—Sensing tongue movements enables various applications in hands-free interaction and alternative communication. We propose BARTON, a *BAR*ometer based low-power and robust *TON*gue movement sensing system. Using a low sampling rate of below 50 Hz, and only extracting simple temporal features from in-ear pressure signals, we demonstrate that it is plausible to distinguish important tongue gestures (left, right, forward) at low power consumption. We prototype BARTON with commodity earpieces integrated with COTS barometers for in-ear pressure sensing and an ARM micro-controller for signal processing. Evaluations show that BARTON yields 94% classification accuracy and 8.4 mW power consumption, which achieves comparable accuracy, but consumes 44 times lower energy than the state-of-the-art microphone-based solutions. BARTON is also robust to head movements and operates with music played directly from earphones.**

***Keywords*-Human computer interaction; Ubiquitous computing; Pressure sensors;**

I. Introduction

Sensing tongue movements has attracted increasing research interests due to its applications in human-computer interaction and alternative communication. Fine-grained tongue pose tracking [1] [2] provides Silent Speech Interfaces (SSIs) for speech-impaired patients or high-noise environments *e.g.*, fire-fighting to communicate. Recognition of pre-defined tongue gestures [3] [4] enables hands-free interaction for device control and input in scenarios where we prefer not to interact with physical devices in hand, such as in crowded metros or riding bikes.

Previous research on tongue movement sensing varies in sensing modalities and sensor placement. Tongue-mounted magnetic senors [2] are effective in tongue pose tracking yet intrusive to users. Jaw-attached electromyography (EMG) electrodes [5], cheek-attached resistive textile sensor arrays [5], and radar integrated in helmets [6] are non-invasive, but are cumbersome and uncomfortable to wear. Computer vision based approaches [7] require no sensors worn by users, but only function when the tongue is outside the mouth.

Since it is socially acceptable to wear earphones in daily life and people tend to wear them for extended periods of time, a promising alternative is to monitor tongue movements via in-ear sensing. Pioneer work has demonstrated the viability to monitor heart rate [8], localize pairs of teeth clicks [9], and recognize respiratory-related events such as snoring and coughing [10], and tongue gestures [4] using microphones. In-ear microphone based tongue and jaw activity sensing operates by capturing and processing sounds induced by mouth-related motions. However, these solutions have two drawbacks. *(i)* In-ear microphone sensing is vulnerable to various audible interference *e.g.*, music played from earphones. *(ii)* Audio processing involves sophisticated signal processing *e.g.*, Fast Fourier Transforms (FFTs). Techniques to filter speech, music and other everyday audio interference further complicates the processing pipeline. These frequent computation-intensive operations can easily drain the battery of wearable devices.

In this work, we propose BARTON, a *robust* and *low-power* in-ear tongue movement sensing system using commercial off-the-shelf (COTS) barometers. BARTON directly measures the subtle air pressure fluctuations in the ear canal induced by facial muscle movements attached to the tongue. Due to the low power consumption ($< 5\,\mu$A at 1 Hz sampling rate) and low sampling frequencies (≤ 50 Hz) of the barometers BARTON is able to accurately detect tongue movements with lower computational and power efforts compared to microphone based systems. Since a typical barometer is more sensitive to low frequencies, it is naturally resilient to various audio interference such as music and speech. In addition, due to their small form factors, barometers can be discreetly integrated into headphones and earpieces for complete invisibility.

To enable practical tongue movement sensing with in-ear barometers, multiple challenges need to be addressed. *(i)* How to capture and distinguish the subtle pressure patterns of different tongue movements under tight power constraints? *(ii)* How to avoid low-frequency interference such as head movements? *(iii)* How to implement a low-power tongue movement recognition pipeline?

Contributions. BARTON addresses the above challenges by leveraging the flat frequency response of COTS barometers in ultra-low frequencies (Sec. III) and extracting features of pressure signals from the time domain only. It harnesses the difference in correlation between a pair of barometers to distinguish head movements and tongue movements (Sec. IV-B). The carefully selected feature set, the simple yet effective linear classifiers and the low-power micro-controller implementation make BARTON effective and power-efficient in tongue movement recognition. We implement BARTON

978-1-5386-2129-5/17/31.00 ©2017 IEEE
DOI 10.1109/ICPADS.2017.00013

with COTS barometers and integrate it into COTS earphones (Sec. V). Evaluations show that BARTON is able to recognize three primary tongue movements (left/right/forward) with 94% classification accuracy while consuming 8.4 mW power. It is also resilient to interfering activities like head movements. Furthermore, case studies demonstrate that BARTON still achieves high recognition accuracy even with music played directly from the earphones, which is almost impossible in existing microphone-based solutions.

The rest of the paper clarifies each of the above contributions, beginning with a frequency response measurement of COTS barometers, followed by the detailed design, implementation, and evaluation of BARTON.

II. Related Work

BARTON is related to the following research.

Tongue Movement Sensing Systems. Primarily designed for patients and paralyzed people, precise tongue motion tracking systems usually place dedicated sensors directly on the tongue [2]. Less intrusive approaches utilize EMG electrodes [5] or textile pressure matrices [3] attached to the facial skin to capture tongue-related muscle movements. Contactless techniques include wireless sensing [6] and computer vision [7]. But they either require a radar mounted on the shoulder [6] or the tongue to be outside the mouth [7], which is uncomfortable to wear or socially awkward. BARTON does not aim to replace dedicated assisting tools for patients, but rather to provide a low-cost, energy-efficient and user-friendly tongue movement sensing modality for everyday use.

In-ear Sensing of Mouth Activities. Outer ear interfaces (OEIs) has attracted increasing attention for their non-intrusive sensor placement and the effectiveness to detect mouth-related activities. Bedri *et al.* [13] exploit infrared proximity sensors to detect jaw movements. Bitey [9] recognizes sounds of five different pairs of teeth clicks with a bone-conduction microphone. Ren *et al.* [10] monitor respiratory-related events such as snore and cough using smartphone earpieces. BARTON is inspired by in-ear mouth activity sensing with microphones.

Our work is most relevant to [4] [14], which distinguish four tongue gestures by deriving tongue movement ear pressure (TMEP) signals from a microphone. However, most microphones are optimized for human speech, which requires extra voice rejection algorithms for mouth activity sensing. Instead, BARTON utilizes COST barometers to detect ear canal pressure changes, which operates at a frequency range lower than human voice. Hence BARTON naturally avoids the interference of human speech, and is more privacy-preserving. In addition, BARTON is optimized in energy consumption to enable long-term usage.

III. Detecting Tongue Motions with Barometers

This section presents the feasibility of in-ear sensing of tongue movements using COTS barometers.

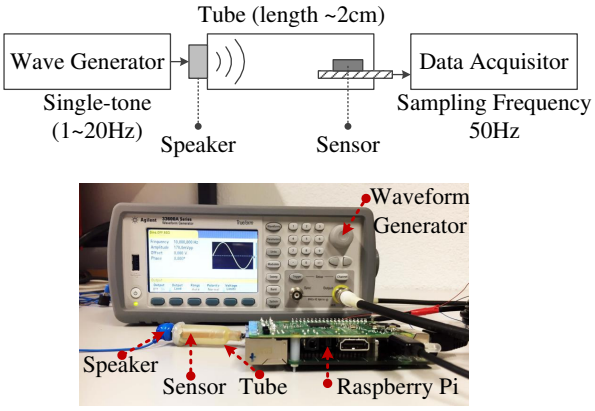


Fig. 1. Frequency response measurement setup.

A. Principles of In-Ear Tongue Movement Sensing

Since the oral cavity is connected to the ear canal, muscle movements of the tongue can induce deformation of, and thus *pressure* changes in the ear canal. Most mouth activities occur within the frequency range of 3.6 Hz to 5.9 Hz [16]. Ear canal pressure variations induced by normal tongue movements also tend to be within the same range. These changes in pressure and airflow generate subtle vibrations and *sound waves* propagating through bones and tissues, which can be acquired by microphones [4], [10]. However, commodity microphones are not optimized for the spectra of non-speech body sounds [15], and customized microphones are often required to obtain high-fidelity sound signals. Instead of detecting sounds of tongue motions using microphones [4], [15], we propose to sense tongue movements by *directly measuring in-ear pressure changes via COTS barometers*.

B. Frequency Response of COTS Barometers

One primary motivation to exploit barometers instead of microphones is low-power, which is important for continuous sensing applications. A COTS barometer consumes around 5 μW for taking a single sample [11]. Digital microphones consume around 1.2 mW but typically only support sampling rates that are too high for our purposes (over 40 kHz). Typical low-cost analog microphones [12] consume around 0.2 mW. However, an analog microphone needs additional signal processing steps such as amplification and analog-to-digital conversion, digital barometers already provide a digital pressure signal. The post-processing steps are usually by a large factor more power hungry than the microphone itself [15].

A natural question arises whether it is feasible to downsample a microphone for low-power tongue movement sensing. We argue it is infeasible because commodity microphones are not optimized for such an ultra-low frequency range.

We validate our claim through the following frequency response measurement of a COTS barometer and a COTS microphone. A desired frequency response in our application demands a *high* and *flat* spectrum in the ultra-low frequency band to facilitate detection of pressure changes even at a

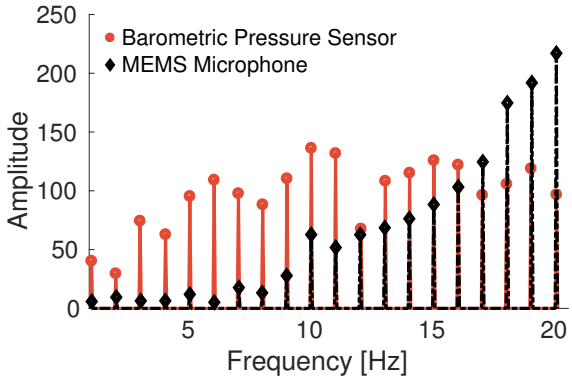


Fig. 2. Results of frequency response tests.

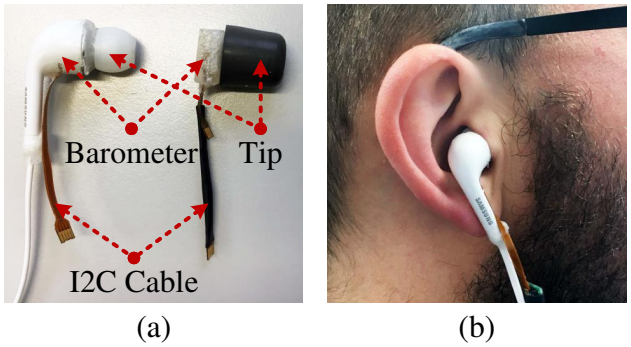


Fig. 3. (a) Sensor sealing for barometric in-ear pressure sensing: custom made with a foam tip (right) and integration into a commercial rubber-tip earpiece (left). (b) Prototype plugged in the ear of a volunteer.

sampling rate below $50\,Hz$. A sampling rate of $50\,Hz$ suffices to capture pressure changes induced by normal tongue movements, which occur below $6\,Hz$ [16]. As will be shown in Sec. VI-B3, even a lower sampling rate of $20\,Hz$ proves to be sufficient for accurate tongue movement sensing.

We used an Agilent 33500A arbitrary waveform generator to create single-tone signals at $\{1, 2, 3, \ldots, 20\,Hz\}$, respectively, to cover the frequency range of normal tongue movements. Note that we also cover a slightly higher frequency range (up to $20\,Hz$) to show the trend in the frequency response of the microphone. The wave generator is connected to a speaker of a headphone, which is placed inside a $2\,cm$ tube (to resemble the ear canal) together with the sensor for testing. The sensor is connected to a Raspberry Pi to collect measurements from the sensor at a fixed rate of $50\,Hz$. Since both the barometer and the microphone are omnidirectional, the orientation of the sensor is irrelevant for the frequency response measurement. Fig. 1 illustrates the setup of the frequency response test. We measure the frequency response for one mainstream MEMS barometer (Bosch BMP280 [11]) and one mainstream MEMS microphone (SPA2410LR5H-B [12]).

Fig. 2 plots the frequency responses of the MEMS barometer and the microphone from $1\,Hz$ to $20\,Hz$. The frequency response of the microphone drops almost linearly in such a low frequency range and cuts off at $10\,Hz$. In contrast, the barometer exhibits consistently flat response and it maintains a moderate magnitude even at frequencies below $10\,Hz$. The results indicate that microphones are usually designed for the audible frequencies ($20\,Hz$ to $20\,kHz$), and are not optimized for such a ultra-low frequency range. Note that ear canal pressure variations induced by normal tongue movements tend to be audible frequencies [16]. Therefore, barometers are better suitable than microphones for capturing in-ear pressure caused by tongue movements.

C. Acquiring Tongue Movement induced Pressure Signals

Previous microphone-based efforts usually need molded earplug housing design [4] or customized foam shells [17] to obtain high-fidelity acoustic measurements. Despite the subtle in-ear pressure signals, we only adopt standard in-ear headphone tips made of foam or rubber (Fig. 3) to place the barometer. Evaluations show that barometer sealing with standard headphone tips is sufficient to shield environmental noise and achieve accurate tongue movement recognition. However, the tips should be of suitable sizes to create an enclosed environment in the ear canal (see Sec. VI-D1). It is also feasible to integrate the barometer into the earpiece to recognize tongue movements (as control commands) while listening to music (evaluated in Sec. VI-D).

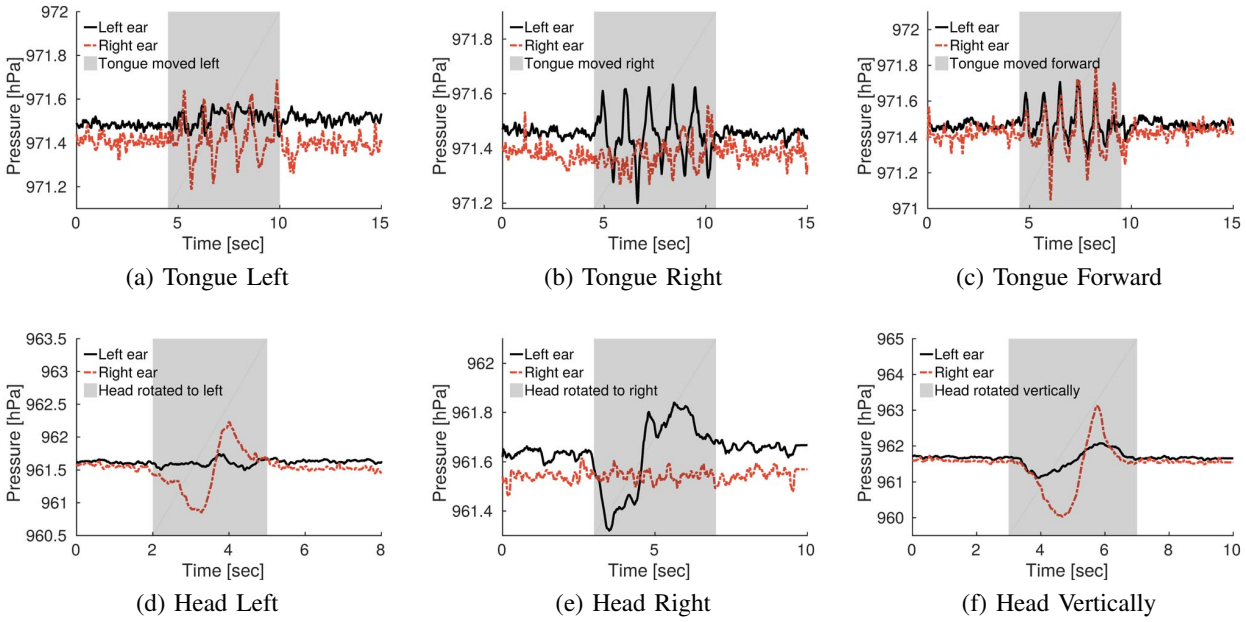
Fig. 4 plots example pressure signals of both ears for three tongue gestures (left, right and forward) and potential interfering activities such as taking the sensors off and head movements. Even though the barometers sample the in-ear pressure signals at only $40\,Hz$, *i.e.*, 50 to 200 times lower than previous works [4] [17], the pressure signals still exhibit distinct characteristics for different tongue gestures and the interfering activities.

IV. Design

This section elaborates on the detailed designs of BARTON including the adoption of barometer pairs, as well as feature selection and classifiers suitable for low-power sensing.

A. Scope

We mainly focus on *primary* tongue movements including *left*, *right* and *forward* [4] [14]. Such a gesture set is sufficient to enable novel hands-free interactions *e.g.*, switching songs through tongue gestures when listening to music using earphones. Complex gestures can be defined by combining the primary tongue movements *e.g.*, *left-right*. The aim of this study is not to cover an exhaustive set of tongue movements, but rather to *(i)* explore *low-power* design with COTS barometers, and *(ii)* investigate *robust* barometer-based tongue movement sensing that works with interfering activities such as head movements and strong ambient noise such as background music from earphones. We envision BARTON as one step further towards always-on tongue interaction interface in free-living environments.



(a) Tongue Left (b) Tongue Right (c) Tongue Forward

(d) Head Left (e) Head Right (f) Head Vertically

Fig. 4. In-ear pressure signals (a) 5 times of moving tongue to the left between 5 to 10 second; (b) 5 times of moving tongue to the right between 5 to 10 second; (c) 5 times of moving tongue forwards between 5 to 10 second; (d) moving head to the left between 2 to 5 second; (e) moving head to the right between 3 to 7 second and (f) moving head vertically between 3 to 7 second.

B. Leveraging Barometer Pair

While Fig. 4 shows notable variations in the pressure signals for different tongue movements, it is challenging to robustly differentiate them under the low-power constraint. *(i)* To save energy, it is prohibitive to apply frequency domain features for classification since they involve intensive power-hungry operations such as Fast Fourier Transforms (FFTs), yet frequency domain features prove to be important in recognizing mouth activities and tongue movements using microphones [4], [15], [18], [19]. *(ii)* The more energy-efficient temporal features alone are insufficient to distinguish tongue gestures (*e.g.*, left and forward, see Fig. 4a and Fig. 4c), and can be easily interfered by head movements (*e.g.*, tongue moving to the left and head moving to the left, see Fig. 4a and Fig. 4d).

To enable robust and low-power tongue movement recognition, we leverage a pair of barometers placed in both ears. Due to the low sampling rate and the low energy consumption of the barometer, the additional energy consumption of an extra barometer is negligible. However, adopting a barometer pair dramatically improves the capability to differentiate tongue gestures, even by using temporal features only. For instance, the pressure signals in both ears are negatively correlated when moving the tongue leftwards (Fig. 4a) yet positively correlated when moving the tongue forwards (Fig. 4c). Head movements usually generate less correlated pressure signals between left and right ear than mouth movements (*e.g.*, Fig. 4f).

C. Features and Classifiers

The raw pressure signals are segmented into windows of samples p_{win} with a window size of 40 samples (1 second) and 50% overlap. All samples p_i are detrended, *i.e.*, $p_i = p_i - mean(p_{win})$, where $mean(p_{win})$ is the mean value within the window. To remain low-power, we exclude frequency-domain features and avoid high-order (*e.g.*, skewness) time-domain features. Table I summarizes a list of candidate time-domain features for tongue movement recognition. The features cover averages, extremes, variances of samples from each individual barometer as well as correlations of pressure signals between the barometers in the left and right ear canals.

To select effective features, we both pick reasonable features and conduct automatic feature selection schemes.

As the adoption of barometer pairs facilitates to differ tongue movements, we select the following features for tongue movement recognition. Specifically, tongue movements generally produce significantly stronger peaks in the pressure signals than head movements (see Fig. 4a and Fig. 4d). Thus it is rational to include features that characterizes extremes such as *min*, *max*, *minDiff*, *maxDiff* and *varDiff*. To distinguish the direction of tongue movement, we select *cov* and *min*/*max*, because *(i)* The pressure signals in the left and right ear canals exhibit strong negative correlations when moving the tongue left and right (see Fig. 4a and Fig. 4b), but notably positive correlations when moving the tongue forward (see Fig. 4c). This observation can be captured by *cov*. *(ii)* The pressure signal shows more notable changes in the right ear when moving the tongue to the left (Fig. 4a), because the muscles on the right are stretched more, and vice-versa. This observation can be captured by *min*/*max*.

Afterwards we adopt an automatic feature selection scheme based on the *Sequential Feature Selection* algorithm [20] to

TABLE I
SUMMARY OF CANDIDATE FEATURES.

Feature	Acronym
Mean of samples within a window	*mean*
Mean of differences of consecutive samples within a window	*meanDiff*
Minimum of samples within a window	*min*
Minimum of differences of consecutive samples within a window	*minDiff*
Maximum of samples within a window	*max*
Maximum of differences of consecutive samples within a window	*maxDiff*
Variance of samples within a window	*var*
Variance of differences of consecutive samples within a window	*varDiff*
Root mean squares of samples within a window	*rms*
Covariance of samples between two ears within the corresponding windows	*cov*
Covariance of differences of consecutive samples between two ears within the corresponding windows	*covDiff*

further optimize the feature set. Finally, we select *min*, *max*, *minDiff*, *maxDiff*, *rms*, *varDiff*, *cov* and *covDiff* for tongue movement recognition.

We use Error Correcting Codes based on binary Support Vector Machines (SVM), K-Nearest Neighbour (KNN) and Decision Tree (DT) as candidate classifiers because they are suitable to be implemented on memory, power and computation limited micro-controllers. We compare the performances of different classifiers in detail in Sec. VI-B1.

V. IMPLEMENTATION

We implement BARTON with COTS barometers as the sensing unit and a micro-controller for low-power tongue movement recognition.

Sensing Unit. We use two Bosch BMP280 barometers [11] to capture pressure signals. The barometers are set to sample at $40\,Hz$ operating in the high resolution mode (*i.e.*, with an internal over-sampling of 8 samples) and internal temperature compensation turned on. This leads to a resolution of 1 Pa with a RMS noise of approximately 1.6 Pa of a pressure sample. The barometers are attached to a rubber or foam cover to be fit into the ear. We also integrate two barometers into a pair of earphones to evaluate the performance of BARTON when playing music from the earphones. The barometers communicate with the embedded processing unit via an I2C bus.

Embedded Processing Unit. We use the launchpad *msp-exp432p401r* featuring the MSP432P401R micro-controller from Texas Instruments, which is based on a Cortex M4 core. It consumes relatively low energy both in active ($240\ \mu A$) and sleep mode ($< 1\ \mu A$). The chip is optimized for floating point arithmetic operations, making it favorable for implementing simple classification algorithms. The micro-controller runs at 3MHz, and the CPU communicates with the barometers over I2C at a speed of 400kHz.

Training is performed offline on a laptop by transferring pressure samples via UART from the barometers. Tongue movement recognition is performed online by using a realtime operating system from Texas Instruments (TI-RTOS) to schedule sampling and classification tasks in the micro-controller.

VI. EVALUATION

In this section, we evaluate the performance of BARTON and conduct case studies to show the effectiveness and robustness of BARTON.

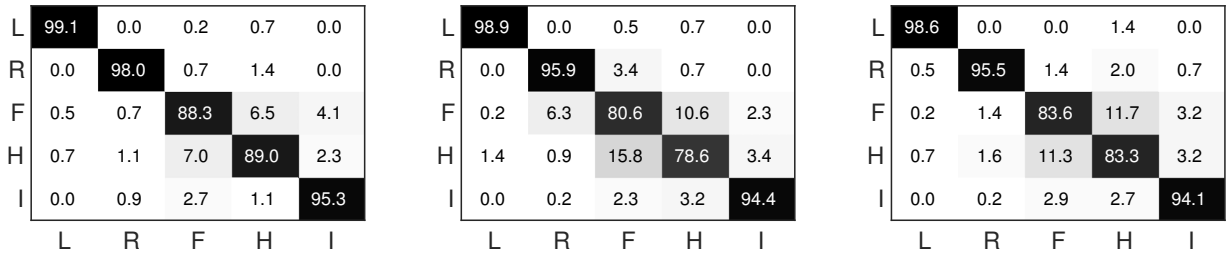
A. Data Collection

We perform measurements of the the three tongue movements and the two interfering activities with one user. The user is consecutively performing each activity multiple times. As mentioned in Sec. IV-C the features from the pressure signals are extracted over a fixed-sized sliding window of 1 sec and 50% overlap. We will discuss the impact of different window sizes and sampling frequencies in Sec. VI-B2 and Sec. VI-B3 respectively. In Sec. VI-D we also collect data from five additional users and investigate the robustness of BARTON across different users.

B. Accuracy

1) Overall Accuracy: Fig. 5 shows the confusion matrices of classifying the three tongue movements (L: Left, R: Right, F: Forward) as well as the interfering activities (H: Head movement, I: Idle) using a sliding window size of 1 second. Each column of the matrix represents the inferred movement, while each row represents the actual movement. The results are based on a 10-fold cross-validation over 444 samples for each activity. The average classification accuracies for all movements (three tongue movements and two categories of interfering activities) are $94\%\pm5\%$, $91\%\pm7.1\%$ and $89\%\pm9.3\%$ for SVM, KNN and decision tree, respectively, while random guesses yield an accuracy of 20% for five gestures. Among the three tongue movements, moving forward is the most likely to be misclassified into head movements. This is expected because moving the tongue forward is characterized by the highly positive correlation between the pair of barometers. Imbalanced stretch of facial muscles and unintentional head movements can incur inconsistent pressure measurements in barometer pair, which leads to misclassification.

2) Impact of Window Size: The choice of window size to extract features affects the classification accuracy and delay. A large window may average out local temporal dynamics while a small window can be prone to noise. Fig. 6 plots the average classification accuracies with different window



(a) SVM (b) KNN (c) DT

Fig. 5. Confusion matrices (accuracy in %) for tongue gesture recognition and different classifiers.

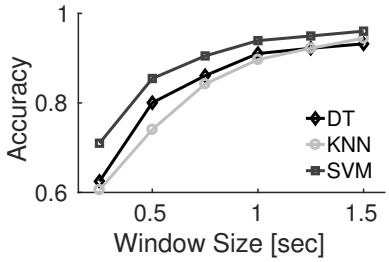


Fig. 6. Impact of window size.

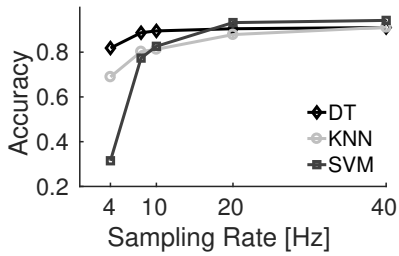


Fig. 7. Impact of sampling rate.

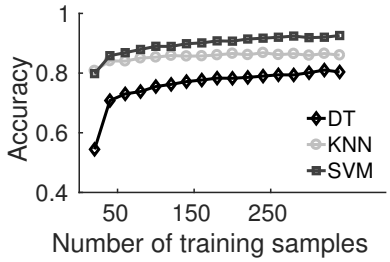


Fig. 8. Impact of amount of training samples.

sizes with a constant sampling rate of 40 Hz. A window size of at least 1 sec achieves good accuracies for all the three classifiers. This results is not surprising because the tongue activities have an average duration of 1 sec while the shortest and longest samples last for 0.75 sec and 1.25 sec respectively. Although the classification slightly improves with a window size above 1 sec it is likely that classifier does detect two individual tongue movements that are performed consecutively within a short time period as a single activity. Consequently the choice of the window size will also greatly depend on the the users preference.

3) Impact of Sampling Rate: The sampling rate for pressure measurements is a trade-off between classification accuracy and energy consumption. A low sampling rate improves power efficiency but the coarse sampled pressure signals may decrease the classification accuracy. Fig. 7 shows the average classification accuracies at different sampling rates. The accuracies remain almost the same if the sampling rate is higher than 20 Hz for all classifiers. An additional interesting observation is the significantly better performance of the Decision Tree compared to the SVM classifier for frequencies below 20 Hz. A possible reason is that the feature loose their distinctiveness with decreasing sampling rate and, thus, the binary classification of the SVM performs worse. In fact, at a sampling rate of 4 Hz the SVM misclassified the majority of the samples as *Idle*. In conclusion using a Decision Tree at low sampling rates helps to save energy while still achieving accuracies above 80%. At high sampling rates the SVM classifier performs best with an accuracy around 94%.

4) Impact of Amount of Training: For practical usage, it is important to minimize the amount of training samples before BARTON is ready to use. We use different amounts of training samples and 100 samples for testing. For each setting we randomly selected the training samples 100 times and calculate the average classification accuracy on the testing samples. Both the training and testing samples are evenly distributed over the five activities. Fig. 8 graphs the tradeoff between the amount of training samples and the average classification accuracy. BARTON requires a minimal of 40 samples for each tongue movement to yield a reasonable classification accuracy of 85% for both the SVM and KNN classifiers. The decision tree needs over 300 samples to reach an accuracy over 80%. It is a well-known challenge to design an optimal decision tree on a small number of training samples [21]. All three classifiers improve their accuracy with increasing number of samples. Particularly the SVM already increases its accuracy to over 90% after 160 samples.

We conclude that it is feasible to train a classifier for BARTON without the need of extensive training data collection.

C. System Performance

Energy efficiency is the primary motivation of our barometer-based tongue movement sensing design. This section evaluates the power consumption and delay of BARTON. Due to the reasonably stable and high accuracy and its simplicity, we adopt the decision tree classifier in the following evaluations. As done before we set the sampling rate to 40 Hz and the classifier uses a sample window of 1 sec.

Fig. 9 shows a complete power trace of BARTON during idle states and performing two classifications. The average

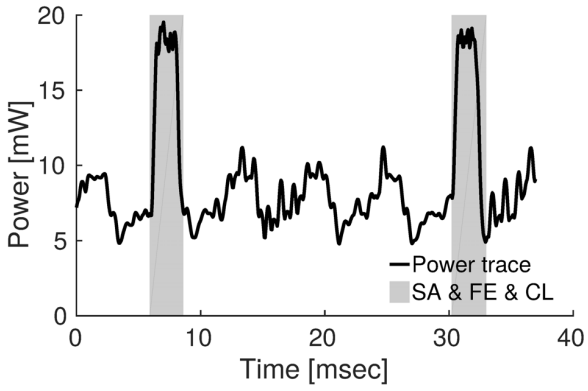


Fig. 9. Power trace.

TABLE II
CLASSIFICATION ACCURACY FOR DIFFERENT USERS WITH INDIVIDUAL TRAINING.

Activity	1 (f)	2 (f)	3 (m)	4 (m)	5 (m)
Left	0.97	0.54	0.91	0.61	0.45
Right	0.9	0.67	0.72	0.5	0.82
Forward	1	0.78	0.63	0.2	1
Head movement	0.93	0.64	0.93	1	0.56
Idle	1	0.97	0.96	1	1
Avg. Accuracy	0.96	0.72	0.87	0.66	0.70

power consumption for a complete cycle (one classification task and idle) is $8.4\,mW$ (about $25\,ms$ in Fig. 9). Specifically, BARTON takes $17.7\,mW$ for sampling (SA), feature extraction (FE) and classification (CL), and consumes $7.6\,mW$ when the CPU of the micro-controller is off. Overall, BARTON is over $44\times$ more energy efficient than the state-of-the-art microphone based implementation [15]. Furthermore, BARTON utilizes $9.1\,\%$ CPU on cycle, which is $2.4\times$ lower than microphone based solutions.

A closer look at the operations of BARTON reveals that on average, BARTON spends $1.53\,ms$ for sampling (SA), $0.14\,ms$ for feature extraction (FE) and $0.1\,ms$ for classification (CL). Depending on the window size, there is a delay of around $0.5\,s$ from the tongue movement till the output of the classification result.

D. Case Studies

In this section, we evaluate the performance of BARTON on different users and its resilience to interference such as music directly played from the earphones.

1) Performance with Different Users: We run the experiments with 5 (User 1 and 2 female, 3 – 5 male) volunteers. Each participant is instructed to put on the BARTON sensing unit himself/herself. All participants are then asked to perform each tongue gesture (left/right/forward) multiple times as well as arbitrary head movement and to remain in idle position. We collect between 12 and 38 samples of each activity and each user. Table II summarizes the average classification accuracy using a 2-fold cross-validation when performing individual training. We observe notable differences in accuracy between

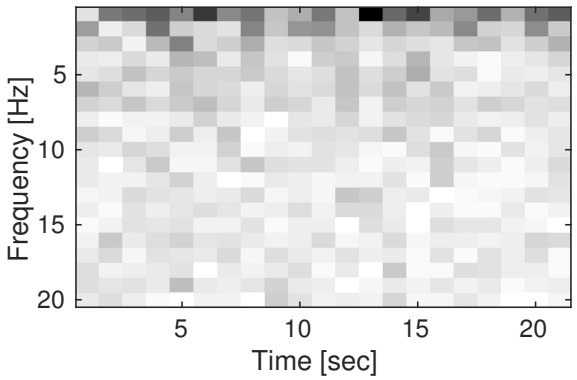


Fig. 10. Spectogram of *Idle* state.

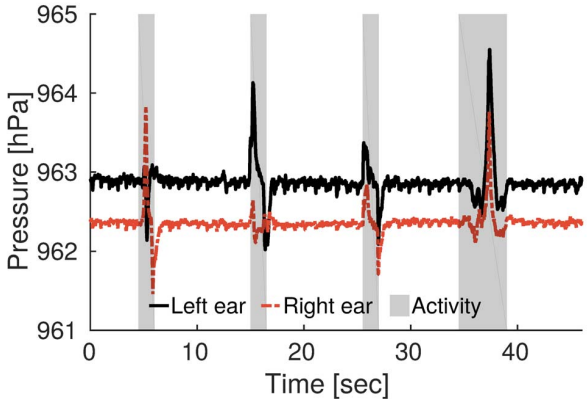


Fig. 11. Pressure signals while playing music.

different users and average accuracies between 0.66 and 0.96. The reasons for the widely varied performances across different users are two-fold.

Feature Differences. The features of the tongue movements are different for each user. For instance BARTON has relatively uniform accuracies for all movements on user 1, while the accuracies for *forward* is extremely low for user 4. The results indicate that user-specific feature sets might be necessary to further improve the classification accuracy. The low accuracy of 0.66 is also related to the second reason.

Headphone Tips. The headphones need to fully enclose the ear canal in order to fully capture the air pressure changes. This was particularly a problem for user 4 and 5. From their feedback, they felt the headphones were not well fit into their ears. As expected, their features are not as distinct as the ones in the training dataset. Although BARTON does not require special earplug housing designs as in [4], it needs headphone tips of suitable sizes to achieve satisfactory classification accuracy.

2) Robustness to Music: In order to facilitate the integration of pressure sensors into commercial in-ear headphones BARTON's performance should not be affected when music

is played. This allows a user to listen to music and controlling a device, *e.g.*, a smartphone, using tongue movements at the same time. We integrate two pressures sensors into commercial rubber-tip headphones, shown in Fig. 3, and conduct measurements while playing music. A *Huwaei P8* smartphone is connected to the headphones and plays different styles of music with the volume set to 75%.

Spectogram. In a first experiment a user is wearing the headphones without performing any tongue activities, *i.e.*, remaining in the idle position. The spectogram in Fig. 10 shows the spectra every 1 sec over a total duration of 20 sec. Between 0 and 10 sec no music is played and between 10 and 20 sec the music is turned on. We observe no notable difference in the spectra between the two phases. This result indicates that the barometers are not sensitive to music played through the headphones.

Tongue Movement Sensing. In a second experiment the user is performing tongue activities while continuously playing music. Starting after 5 sec the user peforms an activity roughly every 10 sec in the order *left*, *right*, *forward* and *head movement*. Fig. 11 shows the pressure signals as well the phases of detected activities. All three classifiers trained in Sec. VI-B1 correctly classified the activities as well as the idle phases between the activities.

These results reveal an important advantage of BARTON over in-ear activity sensing systems based on low-cost microphones. While most of the related works do not consider the interference of music when integrating microphones into headphones, we show that BARTON is not affected by music and works reliably. Due to the high sensitivity of low-cost microphones to frequencies over 100 Hz and the high sampling rates over 1 kHz we suspect that music will affect microphone based systems.

VII. Conclusion

In this work, we propose BARTON, a low-power and robust tongue movement sensing system leveraging COTS barometers. BARTON measures in-ear pressure signals in ultra-low frequencies and extracts time-domain features to differentiate important tongue movements (left/right/forward). We prototype BARTON with COTS barometers, earpieces and a micro-controller. Evaluations show that BARTON achieves comparable accuracy yet consumes 44 times lower energy than the microphone based solutions. BARTON is also robust to head movement and operates even with music played directly from earphones. We envision this work as a promising low-power human-computer interaction mechanism on wearables. In the future, we plan to further improve the robustness of BARTON to diverse activities (*e.g.*, walking and jogging) and investigate user-independent classification.

References

[1] B. Denby, Y. Oussar, G. Dreyfus, and M. Stone, "Prospects for a silent speech interface using ultrasound imaging," in *Proc. IEEE ICASSP*, 2006, pp. 365–368.

[2] H. Sahni, A. Bedri, G. Reyes, P. Thukral, Z. Guo, T. Starner, and M. Ghovanloo, "The tongue and ear interface: a wearable system for silent speech recognition," in *Proc. ACM ISWC*, 2014, pp. 47–54.

[3] J. Cheng, A. Okoso, K. Kunze, N. Henze, A. Schmidt, P. Lukowicz, and K. Kise, "On the tip of my tongue: a non-invasive pressure-based tongue interface," in *Proc. ACM AH*, 2014, pp. 12:1–12:4.

[4] R. Vaidyanathan, B. Chung, L. Gupta, H. Kook, S. Kota, and J. D. West, "Tongue-movement communication and control concept for hands-free human-machine interfaces," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 37, no. 4, pp. 533–546, 2007.

[5] Q. Zhang, S. Gollakota, B. Taskar, and R. P. Rao, "Non-intrusive tongue machine interface," in *Proc. ACM CHI*, 2014, pp. 2555–2558.

[6] Z. Li, R. Robucci, N. Banerjee, and C. Patel, "Tongue-n-cheek: non-contact tongue gesture recognition," in *Proc. ACM IPSN*, 2015, pp. 95–105.

[7] L. Liu, S. Niu, J. Ren, and J. Zhang, "Tongible: A non-contact tongue-based interaction technique," in *Proc. ACM ASSETS*, 2012, pp. 233–234.

[8] S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao, "Musicalheart: A hearty way of listening to music," in *Proc. ACM SenSys*, 2012, pp. 43–56.

[9] D. Ashbrook, C. Tejada, D. Mehta, A. Jiminez, G. Muralitharam, S. Gajendra, and R. Tallents, "Bitey: An exploration of tooth click gestures for hands-free user interface control," in *Proc. ACM MobileHCI*, 2016, pp. 158–169.

[10] Y. Ren, C. Wang, J. Yang, and Y. Chen, "Fine-grained sleep monitoring: Hearing your breathing with smartphones," in *Proc. IEEE INFOCOM*, 2015, pp. 1194–1202.

[11] B. Sensortec, "Bmp280 digital pressure sensor (datasheet)," 2014, accessed Jan 2016.

[12] K. Electronics, "Spa2410lr5h-b low noise zero-height sisonic tm microphone (datasheet)," 2013, accessed Jan 2016.

[13] A. Bedri, D. Byrd, P. Presti, H. Sahni, Z. Gue, and T. Starner, "Stick it in your ear: Building an in-ear jaw movement sensor," in *Proc. ACM UbiComp*, 2015, pp. 1333–1338.

[14] M. Mace, K. Abdullah-Al-Mamun, R. Vaidyanathan, S. Wang, and L. Gupta, "Real-time implementation of a non-invasive tongue-based human-robot interface," in *Proc. IEEE/RSJ IROS*, 2010, pp. 5486–5491.

[15] 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 *Proc. ACM MobiSys*, 2014, pp. 2–13.

[16] S. Karlsson and G. Carlsson, "Characteristics of mandibular masticatory movement in young and elderly dentate subjects," *Journal of Dental Research*, vol. 69, no. 2, pp. 473–476, 1990.

[17] R. Vaidyanathan, M. P. Fargues, R. Serdar Kurcan, L. Gupta, S. Kota, R. D. Quinn, and D. Lin, "A dual mode human-robot teleoperation interface based on airflow in the aural cavity," *International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1205–1223, 2007.

[18] M. Mirtchouk, C. Merck, and S. Kleinberg, "Automated estimation of food type and amount consumed from body-worn audio and motion sensors," in *Proc. ACM UbiComp*, 2016, pp. 451–462.

[19] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," in *Proc. ACM UbiComp*, 2012, pp. 341–350.

[20] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. 100, no. 9, pp. 1100–1103, 1971.

[21] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[22] N. Sissenwine, M. Dubin, and H. Wexler, "The us standard atmosphere, 1962," *Journal of Geophysical Research*, vol. 67, no. 9, pp. 3627–3630, 1962.