Exploring Audio and Kinetic Sensing on Earable Devices

Chulhong Min*, Akhil Mathur*, Fahim Kawsar*†

*Nokia Bell Labs, †TU Delft
{chulhong.min, akhil.mathur, fahim.kawsar}@nokia-bell-labs.com

ABSTRACT

In this paper, we explore audio and kinetic sensing on earable devices with the commercial on-the-shelf form factor. For the study, we prototyped earbud devices with a 6-axis inertial measurement unit and a microphone. We systematically investigate the differential characteristics of the audio and inertial signals to assess their feasibility in human activity recognition. Our results demonstrate that earable devices have a superior signal-to-noise ratio under the influence of motion artefacts and are less susceptible to acoustic environment noise. We then present a set of activity primitives and corresponding signal processing pipelines to showcase the capabilities of earbud devices in converting accelerometer, gyroscope, and audio signals into the targeted human activities with a mean accuracy reaching up to 88% in varying environmental conditions.

CCS CONCEPTS

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

KEYWORDS

Earable, Earbud, Audio sensing, Kinetic sensing

1 INTRODUCTION

The era of *earables* has arrived. Apple fundamentally changed the dynamics of the markets for wireless headphones after it removed the 3.5mm audio jack from iPhone 7. With the explosive growth of the markets, wireless earbuds are also becoming *smarter* by adopting context monitoring capabilities and conversational interface. Recently, sensor-equipped smart earbuds have been actively released into the market, e.g., Apple AirPods, Google Pixel Buds, and Sony Mobile Xperia Ear. Beyond high-quality audio, they are expected to reshape our everyday experiences with new, useful, and exciting services. However, their monitoring capabilities are still limited to the narrow set of exercise-related physical activities.

One of the barriers for modern earables in modelling richer and wider human activities is limited understanding of audio and kinetic sensing on COTS-formed earable devices. Over the last decade, there have been active research efforts to leverage earable sensing, e.g. for eating detection [2], sleep monitoring [14], and energy expenditure monitoring [8]. However, they aimed at detecting specific activities and were mostly evaluated with bulky hardware prototypes. Recently, sensor-equipped earbuds are commercially released, but the research on such COTS-formed devices is limited due to their inaccessible application programming interfaces (APIs).

In this paper, we explore audio and kinetic sensing on COTS-formed earbud devices for a range of human-sensing tasks in the wild. To this end, we prototyped an in-ear wearable instrumented with a microphone, a 6-axis inertial measurement unit (IMU), and dual-mode Bluetooth and Bluetooth Low Energy (BLE). We designed the hardware prototype in an aesthetically pleasing, and ergonomically comfortable form factor (See Figure 1).

With this pre-production and a working prototype, we systematically explore the differential characteristics of the inertial and audio signals in various experimental settings. We looked at how earables compare against a smartphone and a smartwatch concerning some key factors that impact activity recognition pipelines, including signal to noise ratio, placement invariance, and sensitivity to motion artefacts. Analysis of these experimental results suggests that earable sensing is robust in modelling these signals and in most conditions demonstrates superior performance concerning signal stability and noise sensitivities. Inspired by these characteristics, we then design a set of human activity primitives. Activity classifiers are then trained to model these activities with audio and motion data. Early experimental results show that earable sensing can reach up to 88% detection accuracy of targeted human activities.

2 RELATED WORK

Earbuds have been primarily used for a variety of health monitoring applications. LeBoeuf et al. [8] developed a miniaturised optomechanical earbud sensor to estimate oxygen consumption and measure blood flow information during daily life activities. In a recent work, Bedri et al. [2] presented EarBit, a multisensory system for detecting eating episodes in unconstrained environments. The EarBit prototype consisted of two IMUs, a proximity sensor, and a microphone, and it detected chewing events in-the-wild. The authors also presented an insightful analysis on the choice and placement of sensors inside the prototype device. Amft et al. [1] placed a microphone inside the ear canal to detect eating activities and to classify them into four food types. While these works have looked at using ear-worn devices for detecting specific activities with bulky hardware prototypes, the goal of this work is to provide a broad understanding of audio and kinetic sensing on COTS-formed in-ear earbuds for a range of human activities.

Audio and kinetic sensing has been used extensively for detecting human activities and context. Dong et al. [3] used wrist-worn 6-axis IMU to detect eating episodes. Thomaz et al. [17] used a smartwatch

WearSys'18, June 10, 2018, Munich, Germany

ACM ISBN 978-1-4503-5842-2/18/06…$15.00
https://doi.org/http://dx.doi.org/10.1145/3211960.3211970

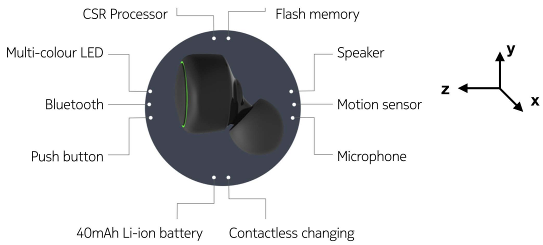

Figure 1: Prototype design of earbuds.

accelerometer to detect the motion of bringing food to the mouth. Hammerla et al. [5] proposed several deep learning models to detect physical activities using wearables. Acoustic sensing approaches have been used to infer human activities and states such as eating [1] and coughing [7]. In this work, we aim at systematically exploring these activities using COTS-formed earbud devices. To this end, we conducted a comparative study of the signal characteristics and detection performance on earable devices and commodity smartphones and smartwatches.

3 HARDWARE PROTOTYPE

For the study, we prototyped earbud devices with multiple sensing modalities. One of the key design goals is to leverage an established form while uncovering opportunities for multi-sensory experience design. As such, we have chosen to augment an emerging True Wireless Stereo earpiece primarily used for hands-free audio experiences - music playback, seamless conversation, etc. This design choice demands that the critical functional attributes are kept as original as possible without compromising their operational behaviour. Our earbud prototype first and foremost is a comfortable earpiece capable of producing high definition wireless audio experience in a compelling form. The earbud is instrumented with a microphone, a 6-axis inertial measurement unit, Bluetooth, and BLE and powered by a CSR processor. Figure 1 shows its final form. Each earbud equips 40mAh battery capacity, weights 20 gram, and has the physical dimension of $18mm \times 18mm \times 20mm$ including the enclosure.

4 UNDERSTANDING SIGNAL BEHAVIOUR

We investigate the opportunities and challenges in human sensing using our earbud prototype. Over the last decade, there have been extensive research efforts on activity monitoring using wearable devices. However, real-world trials of these research efforts have mainly focused on commodity smartphones or smartwatches. The existing approaches may not work well for earable devices as they have different characteristics regarding sensor hardware, position and orientation, and most importantly, the signal patterns.

In this section, we profile the sensing characteristics and discuss the similarities and differences in the audio data and inertial sensor data, across earbuds, a smartphone, and a smartwatch. Based on the analysis, we provide data-driven insights for researchers and developers working on earable sensing.

For the analysis, we used Nexus 5X and LG Watch Urbane for the smartphone and smartwatch, respectively, both which were released in 2015. It is important to note that we do not generalise the values obtained from the study because the performance varies even between smartphones or smartwatches due to their heterogeneity

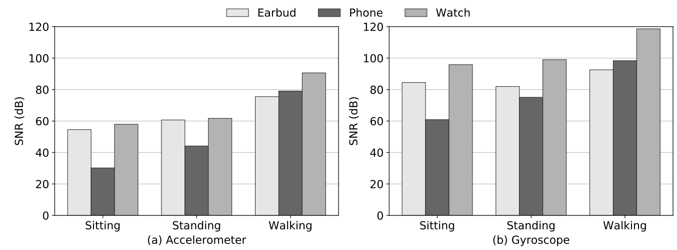

Figure 2: SNR of inertial data across multiple devices.

[13, 15]. Instead, we primarily aim at uncovering the differential characteristics in sensor signals induced by different devices and their positions and orientations on the human body.

4.1 Characterising Inertial Signals

We start by looking at the behaviour of the inertial signals of the earbud and in particular, we study the signal-to-noise ratio and signal sensitivity as a function of placement in two controlled experiments.

4.1.1 Understanding Signal-to-Noise Ratio.

Objective: Inertial sensor data obtained from earbuds, a smartphone, and a smartwatch are directly influenced by the movement of corresponding body parts. The objective of this experiment is to understand how IMU signals behave as it is placed at a different part of our body, e.g., in an ear, on a wrist, etc.

Experimental setup: To quantify the impact of the placement on activity monitoring, we collected the inertial sensor data from earbuds, a smartphone, and a smartwatch, simultaneously. We recruited ten users, and they performed each activity (sitting, standing, and walking) for one minute. Then, we compared the signal-to-noise ratio (SNR) values across the devices.

Results and implications: As shown in Figure 2, earbuds and a smartwatch have higher SNR compared to a smartphone, 20-30dB higher; higher SNR represents less noisy signals, i.e., higher variation compared to the stationary situation. It is reasonable considering that the arms and head have a relatively higher level of freedom of movement, compared to the thigh. For example, the arm will move more than the whole body when a user is walking as people usually swing their arms at the same time. Also, even in the stationary situations, users can move their head and arm to gaze nearby colleagues and type the keyboard, respectively. We omit the result of the gyroscope data as it shows similar patterns.

We further investigate the characteristics of raw data from inertial sensors. Figure 3 shows accelerometer and gyroscope data over time while a user is walking. We can observe that the inertial sensor data on all devices show repetitive patterns to some extent, which are made by walking steps. However, it is also clearly illustrated that the unusual patterns are intermittently observed on an earbud and a smartwatch, e.g., from 10 to 15 seconds. Interestingly, the acceleration data from X-axis (facing direction) on the earbud shows less noise, compared to other axes. We can leverage the benefit of the fixed position of earbuds to remove the noise patterns. While the smartphone shows the least unusual pattern, it does not mean that the smartphone outperforms other devices for physical activity monitoring. In real-life situations, the position and orientation of the

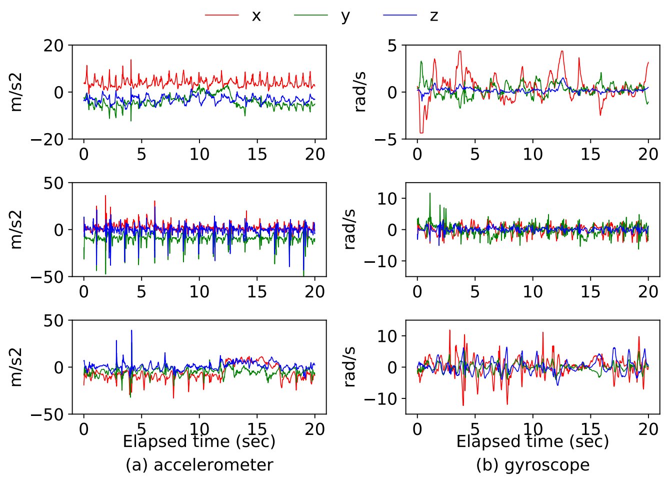

Figure 3: Inertial sensor data on walking (Top: earbud, middle: smartphone, bottom: smartwatch).

smartphone are quite diverse, e.g., in a pocket of a shirt, in a bag, or on a desk.

4.1.2 Understanding Signal Robustness.

Objective: In this experiment, we investigate the signal robustness of the devices with respect to their placements.

Experimental setup: We measure the inertial sensor data from five users with three iterations. Each iteration, the users took off and naturally wore the devices as they usually did. Then, they stood up for 30 seconds without moving. For the smartphone, we asked the users to put in a pocket of pants in the same direction.

Results and implications: To quantify the similarity, we measured the Euclidean distance of signals between two iterations of a single user and calculated the average distance for three cases, ${}_3C_2$. The results show that the average distance between users of the earbud, smartphone, and smartwatch is 0.54, 4.84, and 0.74, respectively; higher distance means more different signals between iterations. Although the smartphone was put in the same direction, its distance was much higher than other devices. This is because the pocket size is relatively larger than the smartphone and thus the smartphone can be placed in an arbitrary direction. On the contrary, we can see that the earbud and smartwatch were placed in a relatively fixed position. The results imply that those devices are more robust to wearing positions, and can be expected to produce consistent accuracy over time and to utilise their absolute direction.

4.2 Characterising Audio Signals

The built-in microphone in earbuds enables a variety of audio sensing tasks such as speech recognition, keyword spotting, and acoustic scene detection. Moreover, its in-ear placement ensures that there is a nearly constant distance between the audio source (speaker's mouth) and audio sensing device (earbud). It reduces the variability in audio data and potentially makes it an ideal device for audio sensing tasks, compared to smartphones and smartwatches.

We explore two key questions about audio sensing capabilities of earbuds with two controlled experiments: 1) how do motion artefacts impact the audio signal characteristics of the earbud, 2) how does the distance of audio source affect the acoustic signal recorded by the earbud.

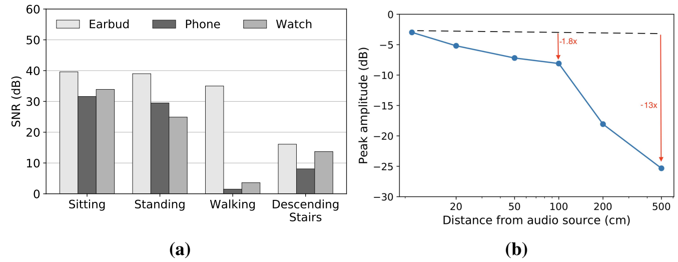

Figure 4: (a) Effect of artefacts. (b) Effect of distance.

4.2.1 Understanding Sensitivity to Motion Artefacts.

Objective: We profile the effect of motion artefacts on the audio data recorded from the earbud and compare it against data from smartphones and smartwatches. More specifically, we hypothesise that motion artefacts, e.g., walking and running, will induce noise in audio signals at a different scale for different types of devices. As such, we compare and report SNR of the audio data collected from multiple smart devices.

Experimental setup: Five users were recruited, and audio SNR was computed in four different physical activities, namely *sitting, standing, walking, stepping down*. To compute the noise induced by physical activities, we first asked the users to perform the target physical activity for ten seconds without providing any speech input. Then, users read a paragraph of text for 30 seconds while performing the same activity. The noise and signal profiles from each activity condition are used to calculate the audio SNR.

Results and implications: Figure 4 (a) shows the average SNR for different devices under multiple activities. We observe that, when users are *sitting*, all the devices report the best SNR. This is expected as the distance between the audio source and the devices, is the lowest in this condition. Interestingly, in the *walking* condition, the SNR for smartphones and smartwatches falls drastically because of the acoustic noise induced by the movement of user's feet and hands. The earbud has a significantly better SNR in walking than other devices, suggesting that noise induced by walking does not impact eaerbud's audio signal significantly. Finally, in the *stepping down* activity, the SNR for earbud is the lowest across all conditions, primarily because of lower impact on acoustic noise induced by this activity. While the SNRs for both smartphone and smartwatch are still lower than the earbud, they show considerable improvement over the walking activity.

4.2.2 Understanding Impact of Distance to Source.

Objective: We study the impact of the distance from the audio source on the audio signal captured by an earbud.

Experimental setup: We placed a laptop (an audio source) at a fixed position in a room and varied the distance of the earbud from the laptop. In each distance condition, we played five different audios on the laptop and measured peak amplitude of the audios recorded from the earbud.

Results and implications: Figure 4 (b) shows the averaged peak amplitude in decibel at different distances from the audio source. At a distance of 10cm, the mean peak amplitude was -2.98 dB. As the distance increases to 1m, the peak amplitude falls by 1.8x, and at a distance of 5m, it falls by 13x. This finding has three key implications for audio sensing tasks on the earbud:

- Earbud is primarily suited for speech and short-range audio sensing activities, but may not be suitable for long-range audio sensing, e.g., acoustic scene detection.
- As the signal amplitude falls heavily at a distance of just 5m, this suggests that audio sensing on earbud is less likely to be affected by acoustic noise in the ambient environment.
- Earbud is less susceptible to false acoustic triggers. For example, speech from a passerby would be less likely to be accepted by the earbud.

5 UNDERSTANDING HUMAN SENSING ON EARBUD

In this section, we present the capability of the earbud in recognising primitive human activities that are extensively studied in the literature. Our objective here is to provide an early indication of the applicability of the earbud as a sensor host in recognising these activities. We have chosen to look at the following activities due to their simplicity and wide applications.

- **Physical activity:** stationary, walking, stepping up, stepping down
- **Head gesture:** nodding, shaking
- **Conversation:** speaking, no speaking

We mainly aim at understanding the prospect of earable sensing, not at optimising the recognition accuracy. As such we adopted the following methodology in our study: i) for physical activities, we compare the recognition characteristics across the devices, ii) for head gestures we only look at earbale sensing and iii) for modelling, we borrowed well-established features and classifiers from the literature. We expect that the findings and lessons from this study will be the base of future research about context monitoring on earable devices.

5.1 Modelling Inertial Signals

We investigate human-centric sensing using inertial sensors on the earbud platform. Here, we target two context types, *physical activity* and *head gesture*.

5.1.1 Data and Model.

We briefly describe here the dataset and models used to assess the performance of the earbud in modelling physical activity and head gestures.

Data: We recruited ten participants and collected the accelerometer and gyroscope data for a set of activities performed by them from three devices - earbuds, a smartphone (Nexus 5X) and a smartwatch (LG Watch Urbane). In particular, Each participant performed a series of activities mentioned above. For physical activities, they were asked to perform each activity for two minutes. For head gestures, they freely conducted several times within one-minute session for each gesture.

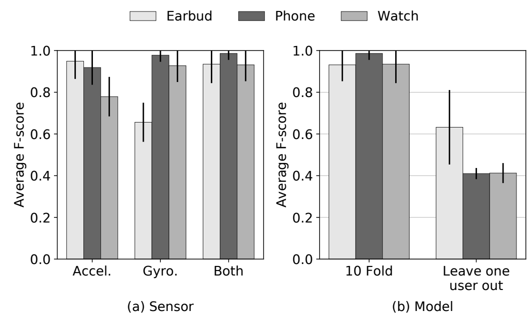

Figure 5: Effect of sensor and model on physical activity detection.

Model: The sampling rates were set to the maximum for each device, 100 Hz for the earbud and SENSOR_DELAY_FASTEST for the rest. We adopt different segmentation approaches for each activity type. For physical activities, we used a time-based sliding window, commonly used for physical activity monitoring. We used 5-second window frame with 95% overlap, i.e., sliding 0.5 seconds. For head gestures, we adopt dynamic segmentation as head gestures are often conducted intermittently in a short time. First, we take two-second window frame with 90% overlap (i.e., sliding 0.2 seconds). Then, we divide each frame into three segments [0.5, 1.0, 0.5 seconds] and compute the variance of the magnitude of the accelerometer data for each segment. We chose the frames only when the variance of the second segment exceeds a threshold, and the variance of the first and last segments does not, i.e., intuitively indicating [*stationary*, *event*, *stationary*].

Various features have been proposed for inertial sensor data, but can be mainly classified into two categories, *time-domain* (mean, median, percentile, RMS, and so on) and *frequency-domain* (spectral energy, information entropy, and so on) features. Here, we used the features reported in [4]. Except for correlation-related features, we take the magnitude of 3-axis sensor values as input. We further applied PCA to reduce the dimensionality of the dataset. We omit the details of feature extraction and PCA methods as their use is not our contribution and already well reported in many works of literature.

We compared the recognition accuracy of 8 popular classifiers, nearest neighbours, linear SVM, RBF SVM, decision tree, random forest, multi-layer perceptron, AdaBoost, and naive Bayes. From our evaluation, the nearest neighbour outperformed the rest of the classifiers. Thus, we report the experimental results using the nearest neighbour. For the analysis, we conducted 10-fold cross-validation by default.

5.1.2 Results.

We first look at the results concerning physical activity monitoring and then report our observation with respect to head gesture detection.

On physical activity: To examine the accuracy dependency on the individual, we compare the F-score with *leave-user-out* validation. Leave-user-out trains the model with all the data except for a specific user and tested the model with the very user. Figure 5 (b)

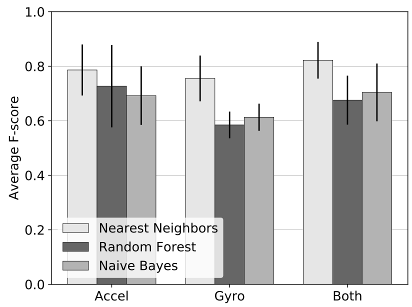

Figure 6: Performance of head gesture detection.

shows the F-score of two validation methods. The F-score of all devices decreases when the leave-user-out validation is used because the model cannot reflect the individual's specific data. However, we can observe that the performance decrease on earbuds is much smaller than that on the smartphone and smartwatch. This is mainly because the wearing position of the earbuds is relatively fixed and stable more than other devices as discussed earlier.

We also investigate the impact of accelerometer and gyroscope on physical activity monitoring. Figure 5 (a) shows the result; *both* represents the case when both accelerometer and gyroscope are used for the feature extraction. While 'both' shows the best F-score regardless of the device, using the accelerometer only also shows comparable accuracy (around 95% on earbuds). It gives the opportunity for energy saving on the earbud platform because the accelerometer is much more energy-efficient than the gyroscope. Interestingly, the gyroscope on earbuds does not contribute to the F-score much, compared to the smartphone and smartwatch. This is mainly because the rotational motion of a head is limited compared to that of an arm and a thigh.

On head gesture: For the evaluation, we collected the ground truth information by manually segmenting the stream of the inertial sensor data. For a fair comparison, we gather all non-gesture data, (non-selected segments in gesture datasets and all segments in activity dataset) and label them to *null* class. Figure 6 shows the F-score while using different sensors and classifiers. The results show that, even with features and classifiers designed for physical activity monitoring, earbuds achieve reasonable accuracy for head gestures. The F-score of the nearest neighbour is 80% when the accelerometer and gyroscope are both used. Even with the accelerometer only, earbuds show 79% of accuracy. We believe we can further optimise the accuracy if we design and use dedicate features and classifiers that well reflect the unique, subtle pattern of head gestures.

To investigate the feasibility of real-time monitoring, we observe the signal patterns made by head gestures. Figure 7 shows the raw data of accelerometer and gyroscope on earbuds when a user was nodding and shaking his head. Each gesture was conducted at the interval of five seconds. We figure out two characteristics. First, the movement of a head when nodding is smaller than the movement when shaking. Second, the different rotational direction of a head is well observed, e.g., there is no much movement in Z-axis direction when nodding, but noticeable movement when shaking. These characteristics enable earbuds to detect head gestures using inertial sensors only.

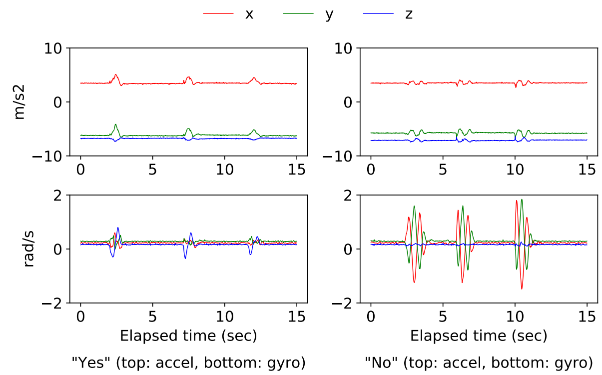

Figure 7: IMU signal behaviour for head gestures.

5.2 Modelling Audio Signals

Detecting with smartphones whether a user is involved in a conversation has been widely studied in the research community [9, 10, 16]. This information provides applications with useful contexts, e.g., smartphone notification delivery could be deferred when a user is involved in a conversation. Here, we focus on the conversation detection using earbuds. As highlighted earlier, our goal is not to demonstrate the state-of-the-art accuracy, but rather to analyse the capabilities of the earbud and explore the possibilities that it creates for future research. We also examine the effect of motion artefacts on the accuracy of the conversation detection. As discussed earlier, motion artefacts add varying amounts of acoustic noise to different devices – as such, we seek to compare the accuracy of conversation detection across multiple devices under different motion conditions.

5.2.1 Data and Model.

We present the dataset and models that we have used to detect the conversational activity with the earbud.

Data: We collected five minutes of data from ten users. Each of them was involved in a separate one-to-one conversation. This process was repeated with three different motion activities: *still, walking, descending stairs*. Also, we collected the data corresponding to the *no conversation* class by gathering two hours of audio data in an office such that no conversation happened within 1m proximity of the sensing devices. Conversations outside this proximate range and other ambient activities were however allowed and were duly recorded by the devices.

Model: We extract 13 MFCC features from the audio data following a sliding window approach (25 ms-long window and overlap of 10 ms). MFCC features from ten consecutive windows are concatenated and used as the feature set for various classifiers. We experimented with four shallow classifiers, namely Random Forests, RBF SVM, naive Bayes, and GMM.

5.2.2 Results.

Figure 8 shows the average F-score of conversation detection using Random Forest; we only report the findings from Random Forest classifier as it outperformed all other classifiers. We observe that the earbud outperforms the smartphone and the smartwatch in the *walking* and *descending stairs* conditions, while its performance (F = 0.88) is close to the smartphone (F = 0.9) in the *still* condition.

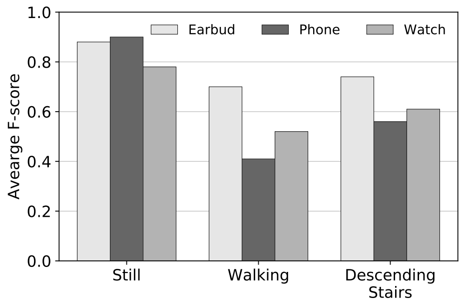

Figure 8: Accuracy of conversation detection.

Particularly in walking, the performance of both smartphone and smartwatch shows a drastic drop (F < 0.5), while the earbud's F-score remains 0.7. This finding is in line with our insights from Section 4, i.e., motion artefacts have a severe impact on smartphones and smartwatches, but a moderate effect on earbuds.

6 APPLICATIONS

Multi-sensory earable devices have the prospect to uncover a variety of applications due to its compelling form, a primary established purpose, and universal acceptability. We briefly discuss a few such applications areas.

Personalised health monitoring: Earables with audio-kinetic models can be effectively used for monitoring a variety of physiological and psychological attributes, e.g., physical activity, diet, head movements, emotion, stress, etc [1–3]. This will help us profoundly achieve medical-grade diagnosis and personalised medicine services with consumer-grade appliances.

Contextual notification: Many recent research works have focused on understanding the receptivity of mobile notifications and predicting opportune moments to deliver notifications in order to optimise metrics such as response time, engagement and emotion. We believe that sensory earables' capabilities in understanding situational context can be incorporated in designing effective notification delivery mechanisms in the future.

Lifelogging: Today's lifelogging applications are primarily vision based. We argue that audio and motion can collectively capture users' moments emphatically and can help them intuitively experience their past.

Social computing: A rich body of literature has focused on tracking face-to-face interactions [6, 9] and the impact of the space on enabling them [11, 12], however often in a constrained setting. Sensory earables will uncover avenues for conducting such studies at scale for new insights.

7 CONCLUSION

We explored audio and kinetic sensing on COTS-formed earbud devices. We systematically explored the differential characteristics of the audio and inertial signals across earbuds, a smartphone, and a smartwatch and showed the capability of earbud devices as a robust platform for human sensing. We showed that earable device achieves a mean accuracy up to 88% in varying conditions. Our results highlight the enormous opportunities for designing multi-sensory applications with earbles.

REFERENCES

[1] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of chewing sounds for dietary monitoring. In *International Conference on Ubiquitous Computing*. Springer, 56–72.

[2] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 37.

[3] Yujie Dong, Jenna Scisco, Mike Wilson, Eric Muth, and Adam Hoover. 2014. Detecting periods of eating during free-living by tracking wrist motion. *IEEE journal of biomedical and health informatics* 18, 4 (2014), 1253–1260.

[4] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and Joao MP Cardoso. 2010. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing* 14, 7 (2010), 645–662.

[5] Nils Y Hammerla, Shane Halloran, and Thomas Ploetz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880* (2016).

[6] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: family-driven mobile intervention care for children with language delay. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1283–1296.

[7] Eric C Larson, TienJui Lee, Sean Liu, Margaret Rosenfeld, and Shwetak N Patel. 2011. Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 375–384.

[8] Steven F LeBoeuf, Michael E Aumer, William E Kraus, Johanna L Johnson, and Brian Duscha. 2014. Earbud-based sensor for the assessment of energy expenditure, heart rate, and vo2max. *Medicine and science in sports and exercise* 46, 5 (2014), 1046.

[9] Youngki Lee, Chulhong Min, Chanyou Hwang, Jaeung Lee, Inseok Hwang, Younghyun Ju, Chungkuk Yoo, Miri Moon, Uichin Lee, and Junehwa Song. 2013. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 375–388.

[10] Chengwen Luo and Mun Choon Chan. 2013. Socialweaver: Collaborative inference of human conversation networks using smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 20.

[11] Afra Mashhadi, Akhil Mathur, Marc Van den Broeck, Geert Vanderhulst, and Fahim Kawsar. 2016. Understanding the impact of personal feedback on face-to-face interactions in the workplace. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 362–369.

[12] Akhil Mathur, Marc Van den Broeck, Geert Vanderhulst, Afra Mashhadi, and Fahim Kawsar. 2015. Tiny habits in the giant enterprise: understanding the dynamics of a quantified workplace. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 577–588.

[13] Akhil Mathur, Tianlin Zhang, Sourav Bhattacharya, Petar Veličković, Leonid Joffe, Nicholas D. Lane, Fahim Kawsar, and Pietro Lió. 2018. Using Deep Data Augmentation Training to Address Software and Hardware Heterogeneities in Wearable and Smartphone Sensing Devices. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '18)*. IEEE Press, Piscataway, NJ, USA, 200–211.

[14] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, and Tam Vu. 2016. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. ACM, 230–244.

[15] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 2015. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 127–140.

[16] Wai-Tian Tan, Mary Baker, Bowon Lee, and Ramin Samadani. 2013. The sound of silence. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 19.

[17] Edison Thomaz, Irfan Essa, and Gregory D Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1029–1040.