# Challenges with Real-World Smartwatch based Audio Monitoring

Daniyal Liaqat
University of Toronto
Vector Institute

Robert Wu
University Health Network
University of Toronto

Andrea Gershon
University of Toronto
Sunnybrook Health Sciences Centre

Hisham Alshaer
Toronto Rehabilitation Institute

Frank Rudzicz
Toronto Rehabilitation Institute
University of Toronto
Vector Institute

Eyal de Lara
University of Toronto

## ABSTRACT

Audio data from a microphone can be a rich source of information. The speech and audio processing community has explored using audio data to detect emotion, depression, Alzheimer's disease and even children's age, weight and height. The mobile community has looked at using smartphone based audio to detect coughing and other respiratory sounds and help predict students' GPA. However, audio data from these studies tends to be collected in more controlled environments using well placed, high quality microphones or from phone calls. Applying these kinds of analyses to continuous and in-the-wild audio could have tremendous applications, particularly in the context of health monitoring. As part of a health monitoring study, we use smartwatches to collect in-the-wild audio from real patients. In this paper we characterize the quality of the audio data we collected. Our findings include that the smartwatch based audio is good enough to discern speech and respiratory sounds. However, extracting these sounds is difficult because of the wide variety of noise in the signal and current tools perform poorly at dealing with this noise. We also find that the quality of the microphone allows annotators to differentiate the source of speech and coughing, which adds another level of complexity to analyzing this audio.

## CCS CONCEPTS

• **Human-centered computing → Mobile devices**; **Empirical studies in ubiquitous and mobile computing**;

**ACM Reference Format:**
Daniyal Liaqat, Robert Wu, Andrea Gershon, Hisham Alshaer, Frank Rudzicz, and Eyal de Lara. 2018. Challenges with Real-World Smartwatch based Audio Monitoring. In *WearSys'18: 4th ACM Workshop on wearable systems and applications , June 10, 2018, Munich, Germany.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3211960.3211977


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

ACM ISBN 978-1-4503-5842-2/18/06.
https://doi.org/10.1145/3211960.3211977

## 1 INTRODUCTION

Audio data from a microphone can be a rich source of information. The speech and audio processing community has explored using audio data to detect emotion [19], depression [6, 12], Alzheimer's disease [5] and even children's age, weight and height [9]. The mobile community has used audio data from smartphones to detect coughing and other respiratory sounds [8, 13] and predict students' GPA [17, 18]. However, these studies tend to use well placed, high quality microphones and/or more controlled environments. For example, [12] describes the dataset they use as studio quality and [8] uses a smartphone as a neck pendant during a short (few hours) study where participants followed their daily routine. While a shorter, uncontrolled study is more realistic than an in-lab study, the short duration makes it so that participants likely remain cognizant of the device and of being recorded.

Smartwatches, and wearables in general have potential to make continuous and in-the-wild sensing much more feasible. Smartwatches are readily available and come equipped with many different sensors, often including a microphone. Compared to smartphones, which may be in a user's pocket, purse or on a table for portions of the day, a smartwatch is much more likely to be on the user's wrist. This means data from a smartwatch's sensors is more likely to reflect the user's state. Additionally, smartwatches are much easier to use day after day as compared to other types of wearable devices (e.g., a chest belt or sensors embedded in clothing). However, the relatively recent emergence of smartwatches and the difficulty of conducting in-the-wild studies creates much uncertainty. It is unclear what kinds of sounds smartwatches will pick up in in-the-wild environments and whether these sounds will be of high enough quality to enable detection of events of interest, such as speech and coughing.

To answer these questions, we built a system that uses Android Wear smartwatches to record raw audio and other sensor data from patients with chronic lung disease. We recruit patients with chronic lung disease because this work is part of a larger study that uses passive sensor data from smartwatches to monitor these patients. While we focus on a specific population, which may affect some of our numbers, we do not think this affects the generality of our key results. For example, the amount of speech in healthy patients may be higher than in patients with lung disease who have difficulty breathing. However, our finding that we need more robust methods for detecting speech from smartwatch based audio still stands and is relevant to many different applications.

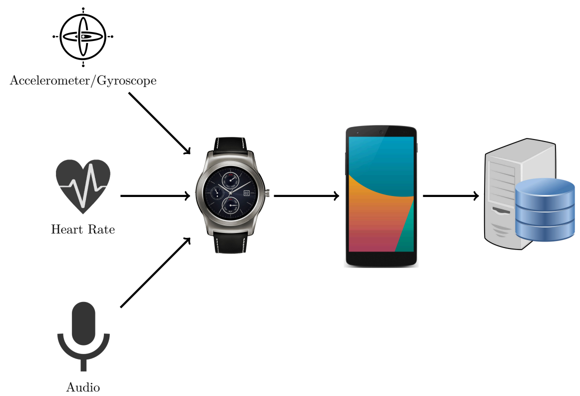


**Figure 1: Flow of data in our data collection system**

To our knowledge, we are the first to record raw, unfiltered audio from an in- the-wild smartwatch. We recruited 16 patients to wear the smartwatch for a three month period while our application recorded data. Our findings include that the audio recorded is of high enough quality to discern speech and respiratory sounds. However, because our data comes from an in-the-wild environment and contains a large variety and amount of noise, algorithms tuned for in-lab studies do not perform well. We find that existing algorithms for Voice Activity Detection (VAD) and cough detection have limited accuracy when applied to our data and that additional sophistication is required to address the challenges of real world audio, which could be an interesting avenue for future research. We also find that a surprisingly high proportion of speech and coughing does not come from the user. These results highlight two problems that will need to be addressed in order for in-the-wild audio analysis to become viable. First, we need more robust methods for automatic event detection such as VAD and cough detection to better handle noisy and inconsistent environments. Secondly, we need reliable methods for distinguishing the source of sounds of interest (user vs. someone else).

## 2 DESIGN

In this section we describe our data collection system and study.

### 2.1 System

Our data collection framework consists of three main components; (1) an Android Wear smartwatch, (2) a phone, and (3) a server. The smartwatch collects sensor data and transmits it to the phone. The phone receives data from the smartwatch and uploads it to a remote server. Finally, the server stores all uploaded data and makes data available for processing and analysis. This data flow is shown in Figure 1.

We use two smartwatch models, the LG Urbane W150 and the Moto 360 2nd Generation, all running Android 6.0.1. For the smartphone, we use either the LG Nexus 5 (Android 6.0.1) or Moto G 3rd Generation (Android 6.0). The smartphone is equipped with a 5GB per month data plan and our data collection framework was tuned to fit within the 5GB per month limit. To prevent users from installing other applications, which could interfere with our data, battery and processing requirements estimations, phones are locked down with a custom launcher and firewall rules.

The smartwatch runs an application that collects sensor data. The main design consideration for this application is battery life. To record sensor data, our application must obtain a partial wake lock from the Android Battery Manager, which prevents the processor from entering sleep mode. Continuously holding this wake lock would drain the battery very quickly. To get around this, we use duty cycling, i.e., recording for a fixed amount of time and then sleeping for a fixed amount of time. Through in-lab testing, we found that for our smartwatches a 20% duty cycling scheme with a 10 minute interval (record for two minutes, sleep for eight) provides enough energy savings to last on average 16 hours on battery, which should be enough to last a full day's use. After deploying we found that these battery saving measures were sufficient with the smartwatches ending 99.9% of days with at least 10% battery remaining.

The smartwatch application has a data collection service that records audio from the microphone as well as data from other sensors. Audio data is sampled at 16 kHz. Unlike regular Android, Android Wear does not support codecs for recording compressed audio. Therefore, we record uncompressed PCM audio and convert it to MP3 using a copy of the LAME MP3 encoder that we cross compile and bundle with our application. Although lossy compression such as MP3 is undesirable, it is necessary to make data transfers feasible and our annotation and automated methods do not suggest that lossy compression is an issue.

Data transfer from the watch to the phone occurs when the watch is placed on charging. The phone application receives sensor data from the smartwatch over Bluetooth and automatically uploads data to a remote server once per day.

It is worth mentioning that for a practical, production-ready application, transmitting raw audio is not required. Ideally, preprocessing on the phone or smartwatch would either extract events of interest or audio features that are transmitted to the remote server rather than raw audio. However, for our research, we need to be able to evaluate the accuracy of preprocessing and to do that, we need the raw audio.

### 2.2 Study Participants

To recruit participants for the study, we approach patients at three different hospitals and ask them to enroll in a 3 month long study. During the study they are asked to wear a smartwatch that passively collects accelerometer, gyroscope, heart rate and audio data. Patients are informed of the study, its goals, and the invasive nature of the data that we are collecting. We also inform users of the security and privacy measures we are required by the ethics committees to take, such as keeping all data and data transmissions encrypted and stored on privately owned and hosted servers. The biggest hurdle in recruiting users is the privacy concerns associated with continuous recording of audio. Despite the privacy concern, we have been able to find patients who agree to participate in the study.

Patients who agree to participate are shown how to use the smartwatch and smartphone. We include features giving patients some control over their data in order to ease some privacy concerns and make it more likely that patients will agree to participate. Patients are able to stop the smartwatch from recording for a short time and on the smartphone selectively listen to and delete recorded audio. They are advised that an optimal way to use the system is to place the smartwatches charging cradle and the smartphone on their bedside table and plug them both in.

## 3 ANALYSIS AND RESULTS

To date, we have collected over 4,100 hours of audio from 15 patients. This data spans over 1059 days with an average of 75 days per patient and 3.9 hours of audio per day. Patients typically put on the smartwatch between 6am and 9am and take it off between 8pm and 10pm. Although the trial was 90 days long, some patients wore the device longer than 90 days due to difficulties scheduling their off-boarding and some ended their trial early but gave us permission to use the data collected so far.

Based on manual annotation, we characterize the audio data in terms of amount of silence, speech and respiratory sounds. But manual annotation of audio is expensive, time consuming and infeasible for a study of even our scale, so for speech and respiratory sounds, we also evaluate how well existing tools can detect these sounds.

### 3.1 Manual Annotation

To annotate audio, we recruit volunteers to listen through the audio and label speech and respiratory sounds. Because we are working with patients and sensitive audio, we have strict restriction on how we store and use this data. For example, we are required to host the data ourselves on servers within our province. These regulations help keep our patients data secure but also mean we cannot crowd-source annotation.

Manual annotation is expensive and time consuming, so we want to maximize the time our annotators spend listening to useful audio. We do this by removing silent portions of the audio. Our duty cycling means that our audio files as recorded are 2 minutes long. We apply a simple silence detection algorithm to these files that first applies an A-Weighting [4] to the audio signal, followed by a low-pass filter and a moving average. The result of the moving average is compared to a preset threshold to determine if the audio segment contains silence. Using this silence detection algorithm, we find that on average 38.3% of the audio data collected from users contains non-silence and the remaining 61.7% is silence. This proportion ranged from a maximum of 59.3% non-silence down to 20.3% across users who participated in the study. We take non-silence segments of audio and stitch them into longer audio files so that annotators are not constantly loading the next file. This mapping of two minute files to non-silent segments to long files for annotation is maintained so that labels created during annotation can be mapped back to the original two minute file.

During our annotation, we also ask annotators to label coughing, speech, throat clearing, sneezing, sniffling, labored breathing, forced expiration and wheezing. Each label consists of a confidence (low, medium, high) and source (patient, 2nd person, TV/radio). The confidence indicates how sure the annotator is that the label is correct. After a bit of practice, annotators are able to learn the patient's voice in order to identify the source of the event of interest. Contextual information is often useful in identifying non-speech events. For example, if the patient is speaking, stops speaking, coughs and then resumes speaking this is an indication that it was the patient coughing. Additionally, over time annotators were able to learn how the patients coughs sound. The two events of interest that we have found the most occurrences of are speech and cough, which is why we focus on these two for our analysis.

*Speech.* To estimate how much of our audio data is speech, we randomly select one week of audio from eight patients from which speech will be analyzed. After removing silence, we are left with an average of 12.20 hours of audio (4.53 hours SD) per user, which annotators listened to and labeled.

We found that overall, 59% of the non-silent audio was speech. Of the speech, 17.66% was from the user, 17.64% was from another person and 54.35% was from TV/radio. While these proportions may vary between users and populations, it does show that a significant portion of speech comes from non-users. This also poses a challenge in using smartwatch based audio as speech from the patient will have to be differentiated from speech coming from other sources.

*Respiratory Sounds.* Detecting respiratory sounds such as coughing is highly relevant to monitoring lung disease and possibly other health conditions. After annotating 53 hours of silence-removed audio across 7 users we discovered 750 coughs, 238 throat clears and 210 other sounds such as labored breathing, sneezing and sniffling. Figure 2 shows the proportion of labels at each confidence level. While the confidence level is a subjective measure, the proportion of labels at each confidence level can serve as a rough approximation of how clear sounds are in the recorded audio and how confident humans are that they can recognize the sounds. Just over 67% of the annotations were made with high confidence, which shows that humans are fairly confident that they can recognize our sounds of interest in smartwatch based audio.

Looking at the source of coughs, we found that 11.4% of the coughs are not from the patient. This was a surprisingly high proportion given that our patients have a chronic lung disease. When trying to monitor coughing, the proportion of coughs coming from other people may be a source of error worth addressing. As mentioned briefly, we collect other sensor data in addition to audio. It is possible that accelerometer and gyroscope data may be helpful in differentiating the source of coughs.

### 3.2 Automatic Detection

We wanted to evaluate how well existing tools for automatic sound classification perform on our real world data. These tools are often developed and evaluated in controlled environments so validating them is essential if they are to be used in health monitoring applications. We look at tools for detecting speech, known as Voice Activity Detection (VAD), and borrow from existing literature to build a cough detection model.

[1] https://webrtc.org

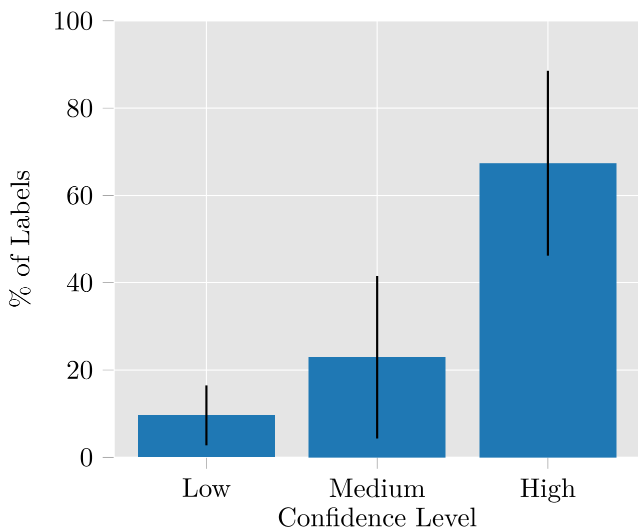


**Figure 2: Proportion of annotations at the three confidence levels**

*Speech.* To evaluate speech detection, we look VAD tools from WebRTC's[1], Loizou [10], Giannakopoulos[2], and LIUM SpkDiarization [11]. WebRTC's VAD has a parameter to control the aggressiveness of the VAD that ranges from 0 to 3, where 0 is the least aggressive about filtering out non-speech and 3 is the most aggressive. The proportion of audio each of these VAD tools classify as speech is shown in Table 1. Interestingly, while speech makes up 59% of the audio, most of these tools were too lenient and classified around 90% of audio as speech (the exception being VAD(2) at 80% and VAD(3) being far too strict at 2%). One explanation for this is that these tools were developed and tested on more consistent audio sources. LIUM for example, was developed for TV and radio broadcasts, [10] used curated dataset of in-lab recordings and [15] assumes that the level of background noise is low.

It is clear that these tools cannot be used as-is on real world smartwatch based audio. Further sophistication is required to not only filter the vast types of noise present in real world audio but also differentiate speech from different sources.

*Respiratory Sounds.* To build an automatic cough detector, we take inspiration from [2], [14] and [1]. We use similar features and machine learning methods as these studies. For feature extraction, we use OpenSMILE [3] to extract spectral features, zero crossing rate, signal energy and Mel-Frequency Cepstral Coefficients from our audio signal using 0.5 second windows with a 0.25 second step. These features, along with annotations from volunteers, are used to train a random forest with an 80/20 random split for training/testing. The average classification accuracy over 100 iterations using monte carlo cross-validation is shown as a confusion matrix in Figure 3.

[2]http://www.mathworks.com/matlabcentral/fileexchange/28826-silence-removal-in-speech-signals

| Method | Speech Proportion (%) |
|---|---|
| WebRTC(0) | 96 |
| WebRTC(1) | 95 |
| WebRTC(2) | 80 |
| WebRTC(3) | 2 |
| Loizou [10] | 91 |
| Giannakopoulos[2] | 92 |
| LIUM | 89 |
| Annotation | 59 |

**Table 1: Proportion of speech in our non-silence audio data as estimated by different tools and from manual annotation.**

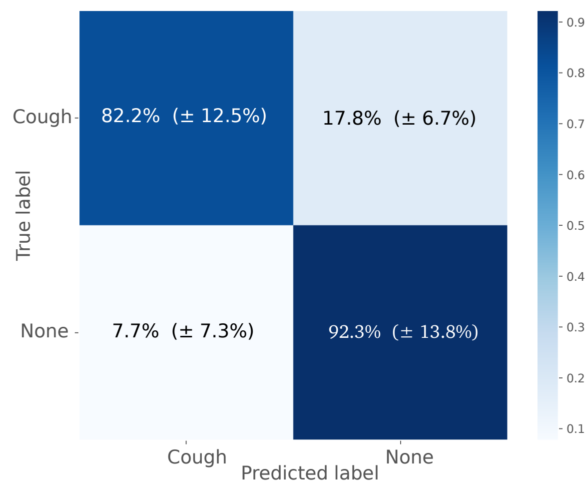


**Figure 3: Confusion matrix for detecting coughing and clearing throat sounds using a Random Forest.**

Our feature selection and classifier is inspired by [2], [14] and [1]. However, our classifier does not perform as well as these previous studies. For example, [1], has a sensitivity of 92.8% and specificity of 97.5% in detecting coughs. Our implementation has a slightly lower sensitivity of 91.43% and significantly lower specificity of 83.83%. Similar to speech, we think this is because these studies use higher quality microphones in more controlled environments. Additionally, in these controlled environments, it is unlikely that there are coughs from other sources so these studies do not attempt to differentiate coughs from the user vs other people.

As discovered through manual annotation, 11.4% of coughs did not come from the user. For the cough classifier, we do not take this into account. Coughs are labeled and classified as coughs regardless of the source. However, for real applications of cough detection, this may be a source of error worth addressing. Additional classifiers could be used to determine whether a given cough came from the user or another person.

From our annotation, it is clear that cough and speech signals are present in the audio. The challenge is the significant amount of other noise present in the signal. This shows again, that more sophisticated methods are required to filter noise and to determine whether the user or someone else is coughing.

## 4 SUMMARY AND DISCUSSION

We find that the quality of audio from the smartwatch is good enough for humans to be able to detect speech and respiratory sounds. Furthermore, the quality is even high enough discern the source of speech and respiratory sounds. We also find that in real-world audio there is a lot of noise which makes automatic classification more challenging. However, we feel that borrowing more from the audio processing community and utilizing advances in machine learning will yield solutions more robust to the kinds of noise seen in real-world audio.

We also saw that simple detection may not be sufficient. We found that 54% of speech in our data was from a TV or radio. In an in-lab, it is unlikely that there would be any sound from a TV. The exact proportions we report are not as important as the fact that there is a significant amount of unexpected sound. The exact proportions can vary based on the population or even location. For example, teenagers who spend a large portion of their day at school may have a lower proportion of TV sounds and higher proportion of "other speaker" sounds. Regardless of the proportion of sounds, for real-world sensing applications, other sources of sound have to be considered and adjusted for. The task of identifying who is speaking when in an audio signal, known as speaker diarization, is a known challenge and will be highly important to wearable audio sensing. However, identifying the source of non-speech sounds, such as coughs, is a novel problem and may be relevant to monitoring various diseases. In a wearable context, the tasks of speaker diarization and non- speech diarization may be able to leverage other available sensors such as the accelerometer and gyroscope to make smarter decisions about the origin of sounds. For example, sudden movement co-occurring with a cough detected in audio could be a strong indicator that the cough was produced by the user.

## 5 RELATED WORK

Many studies have used smartphones for monitoring. Crosscheck [16] for example, uses a smartphone to monitor patients with schizophrenia. They record audio amplitude (not raw audio), accelerometer, location information, application usage and Android's Activity recognition API to track symptoms related to schizophrenia. Using a similar platform, StudentLife [17] monitors student mental health and educational outcomes and SmartGPA [18] predicts student's GPA from smartphone sensor data. These works show that smartphone sensor data can be used for a plethora of monitoring tasks. There are benefits and drawbacks to using smartwatches instead of smartphones. A smartwatch is more likely to be consistently on the users whereas a smartphone may be in a pocket, purse, backpack or table. The location of the device can hinder the interpretation of sensor data. For example, a microphone on the wrist is less likely to be muffled than a microphone in a pocket or backpack. On the downside, smartwatches have less computational power, reduced battery sizes and more limited connectivity. Additionally, while a smartwatch may produce more usable data, some data from a smartphone may be of higher quality. For example, during a phone call users are speaking directly into the smartphone resulting in more speech and less noise.

A study by Kalantarian and Sarrafzadeh [7] uses smartwatches to differentiate eating, chewing and speaking. Their audio is recording in a lab setting and because they are interested in eating events, audio is recorded when subjects are eating and the smartwatch is inches from the subject's mouth. Additionally, their audio is recorded in a lab environment, with noise from a mall edited in after the initial recording. Our study takes place in a completely uncontrolled environment which gives us a better representation of real-world audio.

## 6 CONCLUSION

After deploying a smartwatch based sensing application with real patients, we find that the smartwatch microphone is good enough to pick up speech and respiratory sounds. However, extracting these sounds automatically is difficult because real-world audio contains a wide variety of noise and because a surprising proportion of these sounds do not originate from the patient. We found that existing VAD and cough detection tools have poor accuracy when applied to smartwatch based audio and that more work is needed in filtering out the noise seen in real-world data and to determine whether sounds originate from the user.

## REFERENCES


[1] Yusuf A. Amrulloh, Udantha R. Abeyratne, Vinayak Swarnkar, Rina Triasih, and Amalia Setyati. [n. d.]. Automatic cough segmentation from non-contact sound recordings in pediatric wards. In *Biomedical Signal Processing and Control* (2015-08), Vol. 21. 126–136. https://doi.org/10.1016/j.bspc.2015.05.001

[2] Thomas Drugman, Jerome Urbain, Nathalie Bauwens, Ricardo Chessini, Carlos Valderrama, Patrick Lebecque, and Thierry Dutoit. [n. d.]. Objective Study of Sensor Relevance for Automatic Cough Detection. In *IEEE Journal of Biomedical and Health Informatics* (2013-05), Vol. 17. 699–707. https://doi.org/10.1109/JBHI.2013.2239303

[3] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent Developments in openSMILE, the Munich Open-source Multimedia Feature Extractor. In *Proceedings of the 21st ACM International Conference on Multimedia (MM '13)*. ACM, New York, NY, USA, 835–838. https://doi.org/10.1145/2502081.2502224

[4] Harvey Fletcher and W. A. Munson. [n. d.]. Loudness, Its Definition, Measurement and Calculation. In *The Journal of the Acoustical Society of America* (1933-10), Vol. 5. 82–108. https://doi.org/10.1121/1.1915637

[5] Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2015. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer'ss Disease* 49, 2 (Oct 2015), 407–422. https://doi.org/10.3233/JAD-150520

[6] Kathleen C. Fraser, Frank Rudzicz, and Graeme Hirst. [n. d.]. Detecting late-life depression in Alzheimer's disease through analysis of speech and language. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology* (2016). Association for Computational Linguistics, 1–11. https://doi.org/10.18653/v1/W16-0301

[7] Haik Kalantarian and Majid Sarrafzadeh. 2015. Audio-based detection and evaluation of eating behavior using the smartwatch platform. *Computers in Biology and Medicine* 65 (oct 2015), 1–9. https://doi.org/10.1016/j.compbiomed.2015.07.013

[8] Eric C. Larson, TienJui Lee, Sean Liu, Margaret Rosenfeld, and Shwetak N. Patel. [n. d.]. Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011). ACM, 375–384. http://dl.acm.org/citation.cfm?id=2030163

[9] Jill Fain Lehman and Rita Singh. [n. d.]. Estimation of Children's Physical Characteristics from Their Voices. 1417–1421. https://doi.org/10.21437/Interspeech.2016-146

[10] P. C. Loizou. [n. d.]. Speech Enhancement Based on Perceptually Motivated Bayesian Estimators of the Magnitude Spectrum. 13, 5 ([n. d.]), 857–869. https://doi.org/10.1109/TSA.2005.851929

[11] Mickael Rouvier, Grégor Dupuy, Paul Gay, Elie Khoury, Teva Merlin, and Sylvain Meignier. [n. d.]. An open-source state-of-the-art toolbox for broadcast news diarization. Idiap. https://infoscience.epfl.ch/record/192605/files/Rouvier_Idiap-RR-33-2013.pdf

[12] Brian Stasak, Julien Epps, Nicholas Cummins, and Roland Goecke. [n. d.]. An Investigation of Emotional Speech in Depression Classification. 485–489. https://doi.org/10.21437/Interspeech.2016-867

[13] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao. [n. d.]. SymDetector: Detecting Sound-related Respiratory Symptoms Using Smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015) *(UbiComp '15)*. ACM, 97–108. https://doi.org/10.1145/2750858.2805826

[14] Vinayak Swarnkar, Udantha R. Abeyratne, Anne B. Chang, Yusuf A. Amrulloh, Amalia Setyati, and Rina Triasih. [n. d.]. Automatic Identification of Wet and Dry Cough in Pediatric Patients with Respiratory Diseases. 41, 5 ([n. d.]), 1016–1028. https://doi.org/10.1007/s10439-013-0741-6

[15] Theodoros Giannakopoulos. [n. d.]. Silence removal in speech signals - File Exchange - MATLAB Central. http://www.mathworks.com/matlabcentral/fileexchange/28826-silence-removal-in-speech-signals

[16] Rui Wang, Min S. H. Aung, Saeed Abdullah, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Michael Merrill, Emily A. Scherer, Vincent W. S. Tseng, and Dror Ben-Zeev. [n. d.]. CrossCheck: Toward Passive Sensing and Detection of Mental Health Changes in People with Schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016) *(UbiComp '16)*. ACM, 886–897. https://doi.org/10.1145/2971648.2971740

[17] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. [n. d.]. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2014) *(UbiComp '14)*. ACM, 3–14. https://doi.org/10.1145/2632048.2632054

[18] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T. Campbell. [n. d.]. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (2015). ACM, 295–306. http://dl.acm.org/citation.cfm?id=2804251

[19] Rui Xia and Yang Liu. [n. d.]. DBN-ivector Framework for Acoustic Emotion Recognition. 480–484. https://doi.org/10.21437/Interspeech.2016-488