Comparison of Silence Removal Methods for the Identification of Audio Cough Events

Madison Cohen-McFarlane[1], *IEEE Student Member*, Rafik Goubran [1,2], *IEEE Fellow*, and Frank Knoefel [1,2,3]
madison.mcfarlan@sce.carleton.ca

***Abstract*— Sensing technologies are embedded in our everyday lives. Smart homes typically use an Audio Virtual Assistant (AVA) (e.g. Alexa, Siri, and Google Home) interface that collects sensor information, which can provide security, assist in everyday activities and monitor health related information. One such measure is cough, changes of which can be a marker of worsening conditions for many respiratory diseases. Creating a reliable monitoring system utilizing technology that may already be present in the home (i.e. AVA) may provide an opportunity for early intervention and reductions in the number of long-term hospitalizations. This paper focuses on the optimization of the silence removal and segmentation step in an at home setting with low to moderate background noise to identify cough events. Three commonly used methods (Standard deviation (SD), Short-term Energy (SE), Zero-crossing rate (ZCR)) were compared to manual segmentations. Each method was applied to 209 audio files that were manually verified to contain at least one cough event and the average segmentation accuracy, over segmentation and under segmentation results were compared. The ZCR method had the highest accuracy (89%); however, it completely failed under moderate noise conditions. The SD method had the best combination of accuracy (86%), ability to perform under noisy conditions and low prevalence of over and under segmentation (22% and 15% respectively). Therefore, we recommend using an adaptive approach to silence removal among cough events based on the level of background noise (i.e use the ZCR method when the background noise is low and the SD method when it is higher) prior to implementation of a cough classification system.**

I. INTRODUCTION

Technology is embedded in our everyday lives. With the reduction of hardware size and cost, sensor technologies are providing large scale data representing many aspects of day-to-day life. From sensors integrated in the environment to wearable sensors, there is a growing area of research in providing health information that can support health care professionals or provide health related information to an individual directly. Furthermore, sensing technologies are being combined to extract multiple streams of information. An example of this is the smart home, where a variety of sensors are used to ensure security, monitor activity and provide assistance to the user. Smart home monitoring has recently been applied as a way to support older adults in their own homes, reducing long term hospitalizations and improving Quality of Life (QoL) [1].

[1] Department of Systems and Computer Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada.
[2] Bruyére Research Institute, 43 Bruyére Street, Ottawa, ON, Canada.
[3] Physician at Bruyére Continuing Care and Elisabeth Bruyére Hospital, 43 Bruyére Street, Ottawa, ON, Canada.

To date, smart home monitoring systems aim to monitor a particular type of event using a single sensor or sensor system [1]. Some of the physiological events that have been investigated include; respiratory rate monitoring [2], [3], environmental monitoring [4], [5], activity monitoring [6], mobility monitoring [1], [6], [7], and medical condition monitoring [8].

Harnessing currently available sensors and using them to extract health information, for which they were not designed, is a way to increase the likelihood of user adoption. One aspect of smart home applications is the user interface, typically a speaker and microphone system or Audio Virtual Assistant (AVA) that is used to communicate to the user (e.g. Alexa, Siri, Google Home ... etc.) . In this paper, we describe the preliminary step in the development of a system that utilizes audio signals to monitor health related sounds in the home, specifically cough events.

Cough is a biological protective mechanism that can be related to many respiratory diseases (asthma, allergies, cold, etc.) or other environmental and lifestyle choices (smokers, sedentary people, or athletes, etc) [9]. Changes in cough characteristics can provide an important measure of respiratory changes that may be associated with clinical diagnoses [10]. Identification of such changes may lead to early interventions and reductions in long term hospitalizations. Subjective measurements of the frequency of cough events has been used as a measure of respiratory changes in the past [11], however self report measures can be unreliable [10]. More recently a variety of sensor technologies (cough sounds, electromyography, air flow sensors, etc) have been used to measure both the frequency of cough events and the acoustic characteristics [9], [10]. Many of these audio systems rely on idealized environments or contact microphones [12], [13].

There has been some research into the feasibility of cough detection using portable recording systems that harness smartphone microphones and the on-board CPU [14]. The largest challenge using this system is the restrictions to computational capacity associated with the smart phone and the high tax machine learning techniques have on battery power. Most of the aforementioned systems focus on cough detection and counting, with one system reporting the ability to diagnose pertussis, a respiratory disease common among children, from 38 cough recordings [15] and [16] reporting the ability to differentiate between dry and wet coughing events.

Fig. 1. Cough classification process, with the focus of this paper highlighted (silence removal in blue)

Cough identification and classification follows the general audio signal processing work flow (Fig. 1) [17]. First prepossessing is applied to remove common sources of noise above or below the frequency range of interest [17]. The second step is silence removal, where the audio content of interest is segmented from the overall audio file so that the following steps are not applied to segments that do not contain any relevant content, in this case coughing events [17]. Feature extraction is then applied to all the segments followed by classification methods to categorize the audio content [17]. In this case one could use a classifier to differentiate between coughing events and all other audio events or differentiate between cough types that may be related to a particular respiratory condition.

The result of each step relies heavily on the preceding steps. In order to maximize the system's ability to classify cough events, each processing step needs to be investigated and optimized. Furthermore the processing methods used will vary depending on what kind of audio signal is being investigated. Here we are focusing on the non-speech health related sounds, specifically extracting audio bursts associated with coughing events. This paper focuses on the silence removal step (Fig. 1), sometimes referred to as sound event detection. We are assuming that audio content is recorded at home, where there is low levels of noise. Based on this assumption, we compare three different commonly used silence removal methods; Standard Deviation (SD) used in [15], Short-term energy (SE), and Zero-crossing rate (ZCR). Each silence segmentation method was applied to 209 tagged cough events from the DCASE 2018 database [18] and compared to manual segmentation.

II. DATA PREPARATION

A. Cough Signal Acquisition

As there was no standalone cough database available, the DCASE 2018 challenge dataset was identified to contain an audio class of cough events [18]. The aim of the original dataset was to develop a reliable automatic general-purpose tagging system for audio events. The dataset was created by Freesound.org and Google Research's Machine Perception Team to manually annotate Creative Commons Licensed sounds [19]. There were 242 labeled cough sounds in the DCASE dataset. Each cough sound was manually verified to contain low levels of noise and at least one coughing event, resulting in 209 audio files.

B. Prepossessing

All sound files were 44.1 kHz/16 bit WAV files. As cough events peak around 1.5 kHz all cough sounds were low pass filtered, with a cut off frequency of 4 kHz, to eliminate any high frequency events that may be present [20]. Furthermore, each sound file was downsampled to 20kHz.

C. Manual Annotation

Manual segmentation was performed on all 209 audio files aiming to extract the audio bursts associated with a coughing event from the surrounding low amplitude (silence) periods. The manual segmentation was considered the gold standard and compared to the three silence removal methods.

III. SILENCE REMOVAL

Silence removal relies on identifying periods of the audio signal that do not contain audio events when compared to periods that are associated with cough events, thus simple evaluation methods can be implemented. Three different methods were investigated in order to identify the onset and offset of each period of low audio amplitude, which will be referred to a silence. Each method was applied to frames of each audio signal, based on the average length of a cough segment. Cough events are between 0.2 and 1.0 seconds in duration [21]. More frames at a smaller size provide a more accurate identification of the transition from silence to cough event (or visa versa). Therefore, for each cough event there are at least 16 overlapping segments, which translates to a window size of 250 samples (1.25 ms) with a 50 sample (0.25 ms) overlap. When working with multiple audio classes, it may be appropriate to increase the frame size in order to improve the robustness of the system.

A. Standard Deviation

The SD is a representation of variance of a signal. Periods of high variance are a coarse representation of frequency. Higher SD corresponds to audio bursts and lower standard deviations would be associated with silence. For each frame, the SD was calculated using (1).

$$SD(i) = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}|x_i(n)-\mu_i|^2} \tag{1}$$

Where, μ_i is the mean of the current frame, N is the window size, i is the current frame, and n is the current sample. For each frame, the onset and offset of the cough events were identified using a threshold (T). The threshold was calculated based on the mean and standard deviation (2) of a six second window surrounding the frame that had the lowest SD (SD_{min}), which corresponds to a period of silence.

$$T = mean(SD_{min}) + std(SD_{min}); \tag{2}$$

If a six millisecond segment of silence did not exist, equation (2) would lead to a very high threshold and the silence periods would be missed. To compensate, if the computed threshold was larger than one percent of the maximum SD of the audio file, the threshold was set based on a percentage of the maximum SD (3). Additionally, it was noted that some of the audio files obtained were prefiltered, resulting in exactly zero silence periods. Therefore, if the threshold calculated with (2) was equal to zero, the secondary threshold calculated in (3) was also used.

$$T = 0.01 * max(SD) \tag{3}$$

Fig. 2 presents the raw audio, SD and calculated threshold for a given audio file. If the SD of a frame met the threshold and the preceding frame did not, it was labeled as the onset of an audio burst (green circles). If the SD of a frame met the threshold and the following frame did not, it was labeled as the offset of an audio burst (red squares). The onset and offset of each audio burst were then saved for the segmentation step.

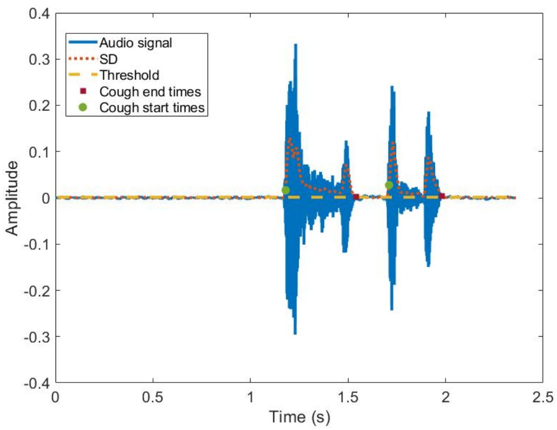

Fig. 2. Audio signal (blue), SD of each frame (orange ':'), threshold (yellow '–'), onset of cough associated audio bursts (green circles), and offset of cough associated audio bursts (red squares) for a given audio file.

B. Short-Term Energy

The SE of an audio signal is a representation of signal power. Each frame is small enough to only contain audio of either high or low power. The SE will be lower for periods of silence than when it is calculated over period of cough related audio bursts (i.e. cough events) [17]. The SE was computed for each frame using (4).

$$SE(i) = \frac{1}{N}\sum_{n=1}^{N} |x_i(n)|^2 \tag{4}$$

Where, N is the length of the frame, i is the current frame, and n is the current sample. For each frame, the onset and offset of the cough events were identified using a threshold based on (2) and (3), where the mean an standard deviation was applied to a six millisecond window surrounding the frame with the lowest SE.

Fig. 3 presents the raw audio, SE and calculated threshold for the same audio file as Fig. 2. If the SE of a frame met the threshold and the preceding frame did not, the frame was labeled as the onset of an audio burst (green circles). If the SE of a frame met the threshold and the following frame did not, the frame was labeled as the offset of the the audio burst (red squares). The resulting data ranges for each audio burst were then saved for the segmentation step.

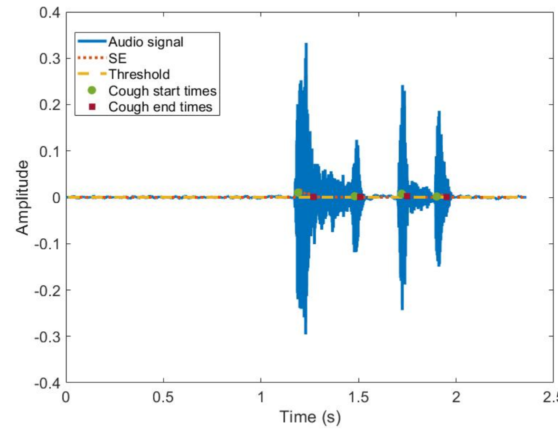

Fig. 3. Audio signal (blue), SE of each frame (orange ':'), threshold (yellow '–'), onset of cough associated audio bursts (green circles), and offset of cough associated audio bursts (red squares) for a given audio file.

C. Zero-Crossing Rate

For each sound file, the times at which the amplitude passes through the zero point was identified. The ZCR is a crude estimate of the frequency of a sound file [17]. Higher numbers of zero-crossings relate to higher frequencies and lower numbers of zero-crossings relate to lower frequencies. As we are looking to differentiate between sound bursts associated with cough events and silence, the cough events would have higher ZCR and thus higher frequencies when compared to the ZCR for adjacent periods of silence. The ZCR was calculated as follows for each frame (5).

$$Z(i) = \frac{1}{2N}\sum_{i=1}^{N} |sign[x_i(n)] - sign[x_i(n-1)]| \tag{5}$$

Where, N is the frame size, n is the sample number, and i is frame number. For each frame, the onset and offset of the cough events were identified using a threshold calculated using (2) and (3), where the mean and standard deviation was applied to a six millisecond window surrounding the frame with the lowest ZCR.

Fig. 4 presents the raw audio, ZCR and calculated threshold for the same audio file as Fig. 2 and Fig. 3. If the ZCR of a frame met the threshold and the preceding frame did not, the frame was labeled as the onset of an audio burst (green circles). If the ZCR of a given frame met the threshold and the following frame did not, the frame was labeled as the offset of an audio burst (red squares). The onset and offset of each audio burst were then saved for the segmentation step.

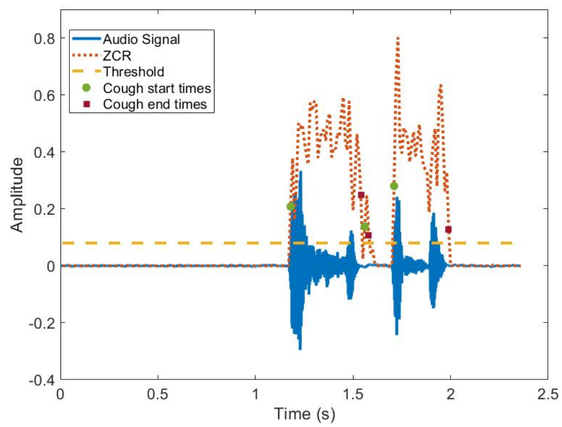

Fig. 4. Audio signal (blue), ZCR of each frame (orange ':'), threshold (yellow '–'), onset of cough associated audio bursts (green circles), and offset of cough associated audio bursts (red squares) for a given audio file.

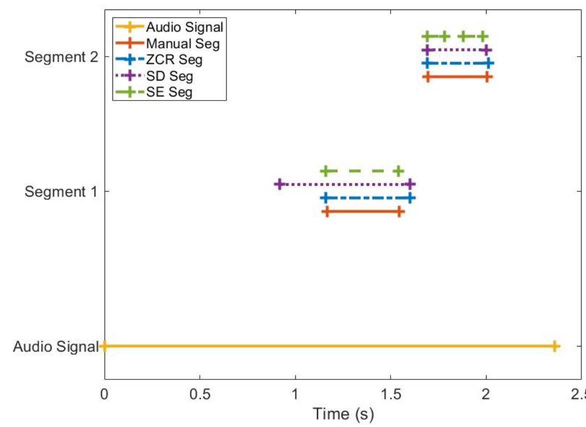

Fig. 5. Onset and offset for the entire audio file (yellow), manual segmentation (orange), ZCR segmentation (blue '-.'), SD segmentation (purple ':') and SE segmentation (green '-') for the current audio file.

IV. SEGMENTATION

To reduce false cough related audio burst identification (Fig. 3 and 4) that may cut off cough events or over segment them, the onset and offset for each method were adjusted prior to final segmentation. Cough events generally contain two consecutive audio bursts connected by a lower amplitude period lasting on average at least 20 milliseconds [21]. Therefore, a single cough audio burst will last at least five milliseconds. Using this, if an identified segment was less than this threshold, it was combined with the previous segment. An example of a too short segment can be seen in the second segment of Fig. 4.

There are instances when the frame identified as either an onset or offset of a cough event did not match the corresponding onset or offset of the manual segment. For instance, the onset of the first segment of Fig. 3 is slightly after the onset of manually labeled cough event. To ensure that the entire cough event was included in the final segmentation one millisecond was added to the beginning and end of all segments.

Finally, some cough events were over segmented, which was common in the SE method and can be seen in the segmentation of both cough events in Fig. 3. Therefore, if two consecutive segments were within five milliseconds of each other, they were combined. The resulting final segmentation onset and offset for each method are summarized in Fig. 5 for the current audio file.

Fig. 5 shows that the ZCR segmentation method performed the closest to the manual segmentation method for this audio file. The small ZCR segment identified originally in Fig. 4 was combined with the first segment based on the segmentation adjustments previously described. The SD method was able to identify the two segments, however the first segment identified starts at an earlier onset than the manual segmentation. The SE method was able to identify the two segments, however the second cough event was over segmented.

The audio signal discussed thus far was a very clean signal with almost zero amplitude present during periods of silence (Fig. 2-4). Fig. 6 shows the segmentation results when these methods were applied to a signal with more background noise (Fig. 7). As can be seen, the ZCR method does not work at all under these circumstances due to the high number of zero-crossings associated with noise. The SD and SE methods are able to identify both segments, however the SE method also identified a segment at the beginning of the file where no true cough event occurred (Fig. 6 and Fig. 7).

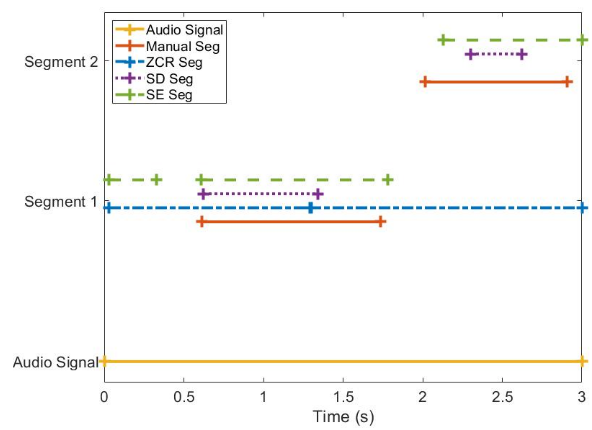

Fig. 6. Start and end times for the entire audio file (yellow), manual segmentation (orange), ZCR segmentation (blue '-.'), SD segmentation (purple ':') and SE segmentation (green '-') for the current audio file.

The three methods were then applied to all 209 audio signals, each containing at least one coughing event and low levels of background noise, from the DCASE dataset [18]. Six metrics were calculated for each audio signal and each silence removal method. The segmentation results from each method (SD, SE, and ZCR) were compared to the manual segmentation results, which were considered the gold standard for this experiment. The average segment

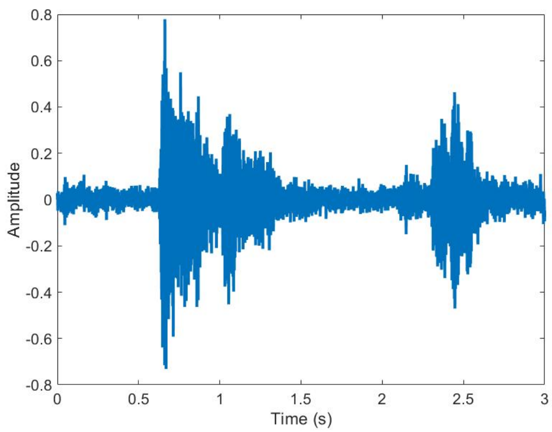

Fig. 7. Audio signal with noisy silence periods corresponding to the segmentation in Fig. 6.

identification accuracy ($Seg_{ID}(\%)$) are reported in Table I for each method. A segment was considered identified if a given method produced a segment that was within $\pm 20ms$ of the manual segmentation range.

If the segment was identified, then it was evaluated in order to determine if it was under segmented (i.e. the given method produced a segment that encompassed at least two manual segments) or over segmented (i.e. the given method produced at least two segments where there was only one manual segment). The average percent of over segmented ($OvSeg_{ID}(\%)$) and under segmented ($UnSeg_{ID}(\%)$) manual segments for each method are also reported in Table I. Finally, for each segment, the average segment onset shift ($On_{shift}(s)$), offset shift ($Off_{shift}(s)$) and duration difference ($Dur_{diff}(s)$) from the manual segmentations were calculated for each method (Table I).

TABLE I
SILENCE REMOVAL SEGMENTATION RESULTS FOR SD,SE AND ZCR

Metric	ZCR	SE	SD
$Seg_{ID}(\%)$	89	83	86
$OvSeg_{ID}(\%)$	31	41	22
$UnSeg_{ID}(\%)$	17	26	15
$On_{shift}(s)$	1.5	1.4	1.3
$Off_{shift}(s)$	1.3	1.2	1.3
$Dur_{diff}(s)$	0.9	1.0	0.9

All silence removal segmentation methods performed reasonably well with the ZCR method presenting the highest segmentation accuracy (89%). When considering the prevalence of over segmentation and under segmentation the results indicate that the SD method had the lowest percent of segments that were both under segmented (15%) or over segmented (22%) when compared to both the ZCR and SE methods. The average shift factors for each method were all very similar ranging from 1.2-1.5 seconds. Finally, segment duration difference when comparing each method to the manual segmentation results showed that the average difference was 0.9 seconds for both ZCR and SD and 1.0 second for the SE method.

V. DISCUSSION

This paper compared three silence removal and segmentation methods. Each method was applied to 209 audio files, each containing at least one cough event and low levels of background noise. The resulting onset and offset for each audio segment were compared to manual annotations for each method (Table, I).

The ZCR method performed the best, however this method fails when considering signals with moderate levels of background noise. As mentioned, some of the original audio files had been preprocessed resulting in silence periods with near zero levels of background noise, which is related to the high performance of the ZCR method. Furthermore, the ZCR method had the second highest prevalence of over segmentation and under segmentation. In contrast, the SD method had the second best silence segmentation accuracy (86%) but had the lowest prevalence of both over and under segmentation. The onset and offset adjustments made prior to segmentation, used to specifically target cough events, had a normalizing effect on the onset shift, offset shift and segment duration differences (Table I).

Depending on the data set under consideration and the effect of under or over segmentation may have on the desired signal an adaptive approach would perform the best. In this case, identification of cough events, over segmentation is a concern as we are dealing with very short audio segments to begin with. If the noise level of the dataset is unknown or varied, the SD method is recommended for silence removal among cough segments, which concurs with the methods used in [21]. If, however it is known that the dataset contains low levels of noise during silence periods, the ZCR method is recommended as it had the highest segment identification accuracy of all three methods. These results correspond to the silence removal method chosen in [15] (i.e. Standard Deviation). Furthermore, the application of this type of silence removal to cough classification problems that did not initially report a silence removal step ([14][16]) may lead to improved classification results.

Future work may include the design and implementation of an adaptive silence removal method. First a rough estimate of background noise level would be computed with both the SD and ZCR methods over the entire audio file. The time-stamp corresponding to the period containing the lowest SD and ZCR would be compared. If there is agreement between these two calculations (i.e. they both identified the same silence period), then the audio file would be classified as having low background noise and the ZCR silence removal method would be employed. If there is no agreement between SD and ZCR segment (i.e. they did not identify the same silence segment) the audio file would be classified as having higher background noise and the SD silence removal method would be implemented.

As mentioned, this paper focuses on the optimization of the silence removal step of the general cough classification process. We focus on the implementation of such a system in a home setting with low levels of background noise. When considering the implementation of this system in a noisy environment (e.g. On a train or in a busy intersection) these methods of silence removal would not be appropriate. To address this issue, future work may also include an investigation into application of higher order audio statistics (commonly used in Voice Activity Detection (VAD)) to identify coughing events in very noisy conditions.

VI. CONCLUSION

The identification, counting and classification of cough events can be a measure of a worsening respiratory condition. Furthermore, implementing methods to do this using sensor technologies that may already be present in the home (e.g. AVA) has the potential to increase user adoption. Here, we focused on the silence removal and segmentation step of the audio processing framework by investigating the performance of three different methods (SD, SE and ZCR). Specifically, the methods described here were tailored to the identification of cough events in low to moderate background noise conditions, thus used threshold parameters chosen based on cough event duration and frequency content. Though the ZCR had the highest performance accuracy (89%), it did not perform well under moderate noisy conditions and had the second highest level of over and under segmentation. In contrast, the SD method preformed well under moderate noisy conditions, had the lowest level of over and under segmentation and had an overall segmentation accuracy of 86%. Therefore, we recommend using an adaptive silence removal method that first evaluates if the background noise is high or low, then implements the appropriate silence removal method (SD and ZCR respectively). Future work will include the implementation this adaptive silence removal method prior to feature extraction and cough related classifications in an at home setting.

ACKNOWLEDGMENT

This work was supported by AGE-WELL NCE Inc., a member of the Networks of Centre of Excellence program, a Government of Canada program that funds partnerships between universities, industry, government and not- for-profit organizations.

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] S. Longhi, A. Freddi, and A. Monteriù, "Human monitoring, smart health and assisted living: Techniques and technologies," in, ser. Healthcare Technologies. Institution of Engineering and Technology, 2017, ch. 1, ISBN: 9781785611506.

[2] F. Benetazzo, A. Freddi, A. Monteriú, and S. Longhi, "Respiratory rate detection algorithm based on rgb-d camera: Theoretical background and experimental results," *Healthcare Technology Letters*, vol. 1, no. 3, pp. 81–86, 2014, ISSN: 2053-3713. DOI: 10.1049/htl.2014.0063.

[3] S. L. Bennett, R. Goubran, and F. Knoefel, "Comparison of motion-based analysis to thermal-based analysis of thermal video in the extraction of respiration patterns," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3835–3839. DOI: 10.1109/EMBC.2017.8037693.

[4] L. Ciabattoni, F. Ferracuti, G. Ippoliti, S. Longhi, and G. Turri, "Iot based indoor personal comfort levels monitoring," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, pp. 125–126. DOI: 10.1109/ICCE.2016.7430548.

[5] S. Rahimi, A. D. Chan, and R. A. Goubran, "Usage monitoring of electrical devices in a smart home," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, IEEE, 2011, pp. 5307–5310.

[6] A. Arcelus, C. L. Herry, R. A. Goubran, F. Knoefel, H. Sveistrup, and M. Bilodeau, "Determination of sit-to-stand transfer duration using bed and floor pressure sequences," *IEEE transactions on biomedical engineering*, vol. 56, no. 10, pp. 2485–2492, 2009.

[7] S. L. Bennett, Z. Ren, R. A. Goubran, K. Rockwood, and F. Knoefel, "In-bed mobility monitoring using pressure sensors.," *IEEE Trans. Instrumentation and Measurement*, vol. 64, no. 8, pp. 2110–2120, 2015.

[8] M. Cohen-McFarlane, J. R. Green, F. Knoefel, and R. Goubran, "Smart monitoring of fluid intake and bladder voiding using pressure sensitive mats," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, IEEE, 2016, pp. 4921–4924.

[9] J. Monge-Alvarez, C. Hoyos-Barcelo, L. M. San Jose-Revuelta, and P. Casaseca-de-la-Higuera, "A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features," *IEEE Transactions on Biomedical Engineering*, 2018.

[10] J. Korpáš, J. Sadloňová, and M. Vrabec, "Analysis of the cough sound: An overview," *Pulmonary pharmacology*, vol. 9, no. 5-6, pp. 261–268, 1996.

[11] S. S. Birring and A. Spinou, "How best to measure cough clinically," *Current opinion in pharmacology*, vol. 22, pp. 37–40, 2015.

[12] J. Monge-Álvarez, C. Hoyos-Barceló, K. Dahal, and P. Casaseca-de-la-Higuera, "Audio-cough event detection based on moment theory," *Applied Acoustics*, vol. 135, pp. 124–135, 2018.

[13] Y. A. Amrulloh, U. R. Abeyratne, V. Swarnkar, R. Triasih, and A. Setyati, "Automatic cough segmentation from non-contact sound recordings in pediatric wards," *Biomedical Signal Processing and Control*, vol. 21, pp. 126–136, 2015.

[14] C. Hoyos-Barceló, J. Monge-Álvarez, M. Z. Shakir, J.-M. Alcaraz-Calero, and P. Casaseca-de-La-Higuera, "Efficient k-nn implementation for real-time detection of cough events in smartphones," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1662–1671, 2018.

[15] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PLOS ONE*, vol. 11, no. 9, pp. 1–20, 2016. DOI: 10.1371/journal.pone.0162128.

[16] H. Chatrzarrin, A. Arcelus, R. Goubran, and F. Knoefel, "Feature extraction for the differentiation of dry and wet cough sounds," in *2011 IEEE International Symposium on Medical Measurements and Applications*, 2011, pp. 162–166. DOI: 10.1109/MeMeA.2011.5966670.

[17] A. Lerch, *An introduction to audio content analysis: Applications in signal processing and music informatics*. Wiley-IEEE Press, 2012, ch. 4.

[18] E. Fonseca, M. Plakal, F. Font, D. P. Ellis, X. Favory, J. Pons, and X. Serra, "General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline," *arXiv preprint arXiv:1807.09902*, 2018.

[19] B. d. J. Alastair Porter and F. Font. (2019). Freesound, [Online]. Available: https://freesound.org/. (accessed: 01.30.2019).

[20] P. Moradshahi, H. Chatrzarrin, and R. Goubran, "Cough sound discrimination in noisy environments using microphone array," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 431–434. DOI: 10.1109/I2MTC.2013.6555454.

[21] A. Murata, Y. Taniguchi, Y. Hashimoto, Y. Kaneko, Y. Takasaki, and S. Kudoh, "Discrimination of productive and non-productive cough by sound analysis," *Internal Medicine*, vol. 37, no. 9, pp. 732–735, 1998.