Towards Detection and Classification Non-Verbal Events and Biosignals in Hearables

Malahat H.K. Mehrban *[1], **Jérémie Voix**†[2,3], **and Rachel E. Bouserhal**‡[1,3]
[1]Electrical Engineering Department, Université du Québec (ÉTS), Montréal, Canada
[2]Mechanical Engineering Department, Université du Québec (ÉTS), Montréal, Canada
[3]ÉTS-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS)

1 Introduction

The miniaturization of sensors coupled with advancements in Machine Learning (ML) have facilitated the achievement of portable health monitoring. In particular, in-ear wearable devices, or hearables, have gained popularity in recent years. This is because within the stable position of the occluded earcanal, hearables can capture various events using only an in-ear microphone [1]. The detection and classification of in-ear biosignals such as heartbeats and respiration have already been achieved with conventional signal processing techniques [2]. However, these conventional techniques rely on peak detection and are affected by movement artifacts. ML methods that are popular for their rapid computation and comprehensive learning capacity may be more robust to such challenging conditions. The goal of this work is to collect an open-access database of various in-ear microphone signals and to develop advanced techniques to detect and classify breathing and heart rate in various acoustically challenging conditions.

2 Background

2.1 Previous Work

Health monitoring methods that use acoustic data to measure physiological signals using traditional digital signal processing (DSP) have already been proposed in the literature. For instance, peak extraction and envelope detection to measure respiration and heart rate from in-ear recordings were proposed in [2]. Authors in [3], used heart sounds recorded from the neck to measure respiration and heart rate using Continuous Wavelet Transform (CWT). In general, DSP methods work well with data captured in controlled environments but are often not robust to real-world disturbances such as noise and movement.

Recently, ML algorithms have been used for various tasks in detection and classification of audio events. Authors in [4], extracted Mel Frequency Cepstral Coefficients (MFCCs) and Spectrogram Image Features (SIF) as features and employed ML algorithms such as Support Vector Machine (SVM), Hidden Markov Models (HMM) and Deep Learning (DL) models like Convolutional Neural Networks (CNNs) to detect sound events in noise-corrupted real-world data. Similarly, in [5], an approach was proposed to detect non-speech audio events using SVM, HMM and Multi Layer Perceptron (MLP). The features which were extracted by the authors to train the aforementioned algorithms consisted of MFCCs, Zero Crossing Rate (ZCR), fundamental frequency, as well as brightness and bandwidth from the extracted spectrogram. Bag-of-Audio-Words (BoAW) or Bag-of-Feature techniques which are inspired by Bag-of-Words to be used for audio event detection were utilized by researchers in [6,7] and [1]. Also, as an efficient ML algorithm to classify non-verbal events in audio data, Gaussian Mixture Models (GMMs) were used in [8], [9], and [10].

Currently, there exists no algorithm capable of detecting and classifying various physiological and non-verbal events captured with an in-ear microphone. In addition, there is a need for a large database of the relevant in-ear microphone signals captured in the various realistic use cases of a hearable. This work will result in a ML algorithm to detect and classify heartbeats and respiration as well as an open-access database of various non-verbal and physiological signals captured with different sensors, including an in-ear microphone.

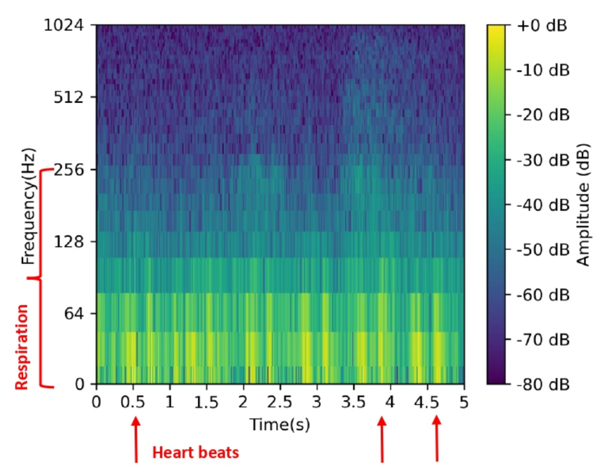

Figure 1: An illustration of the spectrogram of normal nose breathing for one subject is provided, demonstrating that both respiration and heart beat can be captured with the in-ear microphone.

3 Methodology

3.1 Database Creation

To explore the potential of ML for the detection and classification of physiological signals captured from inside the earcanal, more data in various acoustical conditions is required. Data will be collected using an occluding intra-aural earpiece featuring in-ear microphones, outer-ear microphones and miniature loudspeakers in each ear. Recordings

*mmehraban@critias.ca
†jeremie.voix@etsmtl.ca
‡rachel.bouserhal@etsmtl.ca

will be made in protected mode (attenuating outside sounds) and transparent mode (environmental sounds are played back in the earcanal) in quiet as well as in noise. Various human-produced sounds will be also included, ranging from a cough to the blink of an eye.

3.2 Classification of Biosignals

In this section, the steps to achieve the detection algorithm for non-verbal events and physiological signals are described. The primary challenge will be the detection and classification of breathing sounds and heartbeats which are relatively faint and occur simultaneously as other signals. Furthermore, the physiological signal detector and classifier must be compatible with that proposed in [10] which detects and classifies various non-verbal audio events captured from inside the occluded earcanal. Since in-ear recordings have a limited bandwidth, of about 2 kHz, raw audio samples are downsampled fronm 44.1 kHz to 8 kHz. Signals are then framed into 5-second frames. Features are then extracted from each 5-second frame. Various features are of interest including the spectrogram (an example can be seen in Figure 1), MFCCs, ZCR, and Per Channel Energy Normalization (PCEN). Feature reduction are used to reduce the computational complexity of the algorithm. Various classifiers will be explored including CNN and SVM models. Figure 2 illustrated the main structure of this approach.

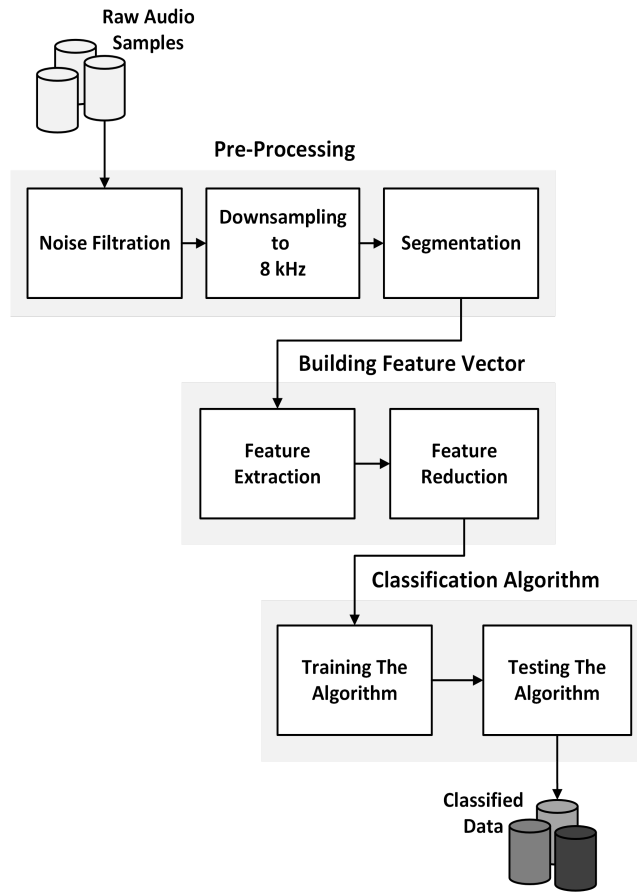

Figure 2: Block diagram of the proposed approach.

4 Conclusions

With the fast development of in-ear health monitoring technologies there is a need for accurate, efficient and quick algorithms to detect and classify real-time information. The proposed open-access database will serve to improve existing ML algorithms and speed up research in this field. In addition, the achievement of the proposed approach will open up the door to various health monitoring applications with hearables, including vital sign tracking as well as emotion classification.

Acknowledgments

This work was supported by ÉTS-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS) and funding from Natural Sciences and Engineering Research Council.

References

[1] Rachel E. Bouserhal, Philippe Chabot, Milton Sarria-Paja, Patrick Cardinal, and Jérémie Voix. Classification of nonverbal human produced audio events: A pilot study. In *Interspeech 2018*, pages 1512–1516. ISCA, 2018-09-02.

[2] A. Martin and J. Voix. In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring. 65(6):1256–1263, 2018-06. PRJ: 209.

[3] Guangwei Chen, Syed Anas Imtiaz, Eduardo Aguilar–Pelaez, and Esther Rodriguez–Villegas. Algorithm for heart rate extraction in a novel wearable acoustic sensor. 2(1):28–33, 2015. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1049/htl.2014.0095.

[4] I. McLoughlin, H. Zhang, Z. Xie, Y. Song, W. Xiao, and H. Phan. Continuous robust sound event classification using time-frequency features and deep learning. 12(9), 2017.

[5] J. Portêlo, M. Bugalho, I. Trancoso, J. Neto, A. Abad, and A. Serralheiro. Non-speech audio event detection. pages 1973–1976, 2009. ISSN: 1520-6149.

[6] Stephanie Pancoast and Murat Akbacak. Bag-of-audio-words approach for multimedia event classification. Technical report, SRI International Menlo Park United States, 2012.

[7] Maximilian Schmitt, Christoph Janott, Vedhas Pandit, Kun Qian, Clemens Heiser, Werner Hemmert, and Björn Schuller. A bag-of-audio-words approach for snore sounds' excitation localisation. page 5.

[8] Jurgen T. Geiger and Karim Helwani. Improving event detection for audio surveillance using gabor filterbank features. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 714–718. IEEE, 2015-08.

[9] Jens Schröder, Jorn Anemuller, and Stefan Goetze. Classification of human cough signals using spectro-temporal gabor filterbank features. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6455–6459, 2016.

[10] Philippe Chabot, Rachel E. Bouserhal, Patrick Cardinal, and Jérémie Voix. Detection and classification of human-produced nonverbal audio events. 171:107643, 2021-01.