Sensor Applications

A Case for an Oral Cavity Based Respiratory Rate Sensor System

Runbei Cheng[1] and Jeroen Bergmann[1]

[1]*Natural Interaction Lab, Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Old Road Campus Research Building, Headington, Oxford, UK, OX3 7DQ*

Manuscript received Aug 31, 2022; revised Sep **, 2022; accepted *** **, 2022. Date of publication *** **, 2022; date of current version *** **, 2022.

Abstract—Respiratory rate has been identified as a promising metric for field-based sports monitoring. While respiratory metrics such as minute ventilation (VE) are commonly used in lab-based metabolic tests, they have yet to be implemented in contact sports which encounter physical impact. This study proposes that breathing can be captured on-field via acoustic sensors embedded inside existing sports gears. Two suitable locations for such a system have been identified, either as a smart mouthguard or an instrumented headgear. The signal-to-noise ratio (SNR) at these placements and their potential for respiratory rate estimation will be compared in this study. Four participants were recruited, and respiratory data were captured in both indoor and outdoor settings. A fast-Fourier transform (FFT) based frequency domain analysis was used to estimate the respiratory rate and determine the breathing rate accuracy. A Wilcoxon signed-rank test was carried out to compare the datasets from the two sensor placements. It was found that the SNR of the oral placement is significantly better than the head placement for both indoor and outdoor settings (indoor: $P = 5.73 \times 10^{-7}, z = 5$; outdoor: $P = 2.12 \times 10^{-7}, z = 5.19$). It was also found that the oral placement had a significantly smaller error compared to the head location when predicting respiratory rate (indoor: $P = 1.72 \times 10^{-6}, z = -4.78$; outdoor: $P = 1.06 \times 10^{-7}, z = 5.32$). In addition, a convolutional neural network (CNN) based classifier was trained to clean up any non-respiratory sounds from recorded audio, which subsequently achieved a 90% test accuracy. This shows is a promising result for demonstrating the viability of a fully automated oral-based respiratory rate monitoring system.

Index Terms—respiratory rate, Sports Monitoring, Smart Mouth-guard, FFT, Machine Learning, CNN.

I. INTRODUCTION

The health and well-being of athletes can be improved via monitoring of their internal loads (physiological responses to physical workloads) [1]. Two methods in particular, survey-based ratings of perceived exertion (RPE) and heart-rate-based metrics, have been dominating the field of on-field internal load monitoring. And while RPE demonstrates a good representation of an athlete's physical exertion, it is difficult to achieve continuous RPE monitoring due to its survey-based nature. It also suffers from reporting biases if RPE outcomes are used in a decision making process. It has been shown that breathing frequency, but not heart-rate, has a strong relationship to RPE [2]. Respiratory rate correlates well to physical exertion [3], and can potentially be used to monitor recovery from exertion better than heart rate [4], thus presenting a promising additional on-field physiological metric to help improve the health and well-being of athletes.

Studies have shown that microphones can be used to obtain respiratory information by applying audio analysis techniques [5], [6]. One study has attempted to implement this idea in an industrial environment by embedding microphones into protective earbuds [6]. A similar approach can be taken in contact sports. Two suitable locations have been identified for microphone placement in contact-sports settings; (i) mounted to a piece of headgear such as a pair of sport safety glasses, a helmet or a scrum cap, or (ii) embedded inside a mouthguard. This study investigates the most optimal microphone placement for measuring respiratory rates in athletes by comparing the signal-to-noise ratio of data captured in these two locations. The performance of a simple frequency-domain-analysis-based respiratory rate estimation algorithm is determined, when applied to the different data sets.

Corresponding author: J.Bergmann (e-mail: jeroen.bergmann@eng.ox.ac.uk).
Associate Editor:
Digital Object Identifier:

II. MICROPHONE LOCATION VALIDATION

This study was conducted under the ethics approval granted by the Medical Sciences Interdivisional Research Ethics Committee (R70833/RE001). Data was simultaneously collected from body-worn sensors that were placed in the oral cavity and on the head of the volunteers. In addition, a third sensor was placed within 30cm away from the participants off person, acting as an external control set. This third location was selected akin to the sensor placement reported by Nam et al 2015 [5], which claimed that breathing sounds can be picked by a consumer-grade microphone placed as far as 30cm from a person. A Signal to noise ratio (SNR) analysis was carried out to compare the quality of data collected in all three locations. A simple fast Fourier transform (FFT) based frequency analysis was used to obtain respiratory rates from all three data sets and their accuracies were compared.

A. Data Acquisition

Four healthy participants, one female and three males, between the ages of 21 to 30 were recruited for this study. Two data capture sessions were performed with each participant individually, one session under a quiet indoor condition, and the other under an outdoor condition simulating a more noisy sports pitch. During each session, the volunteer was fitted with a customized mouthguard integrated with electronics. The mouthguards were thermoformed using the participants' dental impressions. The electronics inside the mouthguards are each equipped with a digital microphone (MP34DT05-A, STMicroelectronics, Switzerland) with a 64 dB

signal-to-noise ratio and -26 dBFS ± 3 dB sensitivity, sampling at 8 kHz and 16 bits per sample. A second set of identical electronics were also attached to goggles, which were given to the participants to represent the headgear placement. And a third off-person set of identical microphones were placed within 30cm of the participants to function as an external control set.

Each data capture session aimed to collect at least 10 minutes of pure breathing audio, which was divided into shorter continuous breathing. During data capture, the participants were instructed to breathe naturally. The beginning of each clip was marked by the participant performing an audio synchronization action consisting of holding their breath, which helped to check synchronisation of the data captured by different microphones. The participants were also asked to log their breathing using a time-stamped key-logging mobile app written in C#. These logs were then used as the ground truth during data analysis. The participants were also instructed to hold their breath at multiple instances throughout each data capture session, to collect normative background noise for the SNR analysis.

B. Data Processing

The raw audio files acquired from the data capture were clipped into multiple clips of continuous breathing, each around one minute in duration. Eighty clips of one-minute-long continuous breathing audio were cropped from the raw data per microphone placement, 40 under indoor conditions and 40 for the outdoor condition. Open-source audio editing software, Audacity® (https://github.com/audacity/audacity) was used during the manual labelling of the data, which helped to label the data accurately down to milliseconds.

SNRs were calculated using the root mean squared values of the pure breathing clips and the average noise level obtained during the requested non-breathing episodes for each participant. The SNR is given by the following formula:

$$SNR = 20 \cdot log(\frac{RMS_{signal}}{RMS_{noise}}) \quad (1)$$

The clips of continuous breathing were then pre-processed through a series of filters. A 4th order Butterworth band-pass filter of 100Hz-400Hz was applied forward and back on the data to get rid of low and high-frequency noises outside of the range of breathing [3]. Following that, the data were downsampled to 1600Hz. Audio envelopes were then extracted via a Hilbert transformation followed by a moving average which subsequently further reduced the data to 8Hz.

A FFT was performed on the sound envelopes to transform the data into the frequency domain. A peak detection algorithm was then implemented to look for a pair of high-power peaks in the FFT separated by a factor of two, with the lower one representing the number of full breath cycles, and the higher one representing the number of inhales and exhales combined (Figure 1). The peak with the lower frequency value was then selected as the respiratory rate of the audio clip. Figure 2 shows a block diagram of the algorithm pipeline. The respiratory rate estimations were compared to the respiratory rate logged by the participants during the data collection to give an algorithm percentage error (PE).

A one-sample Kolmogorov-Smirnov (KS) test was used to determine whether the SNR and the PE were normally distributed. The KS test indicated that the data were not normally distributed, so a non-parametric Friedman test was carried out for both the SNR and

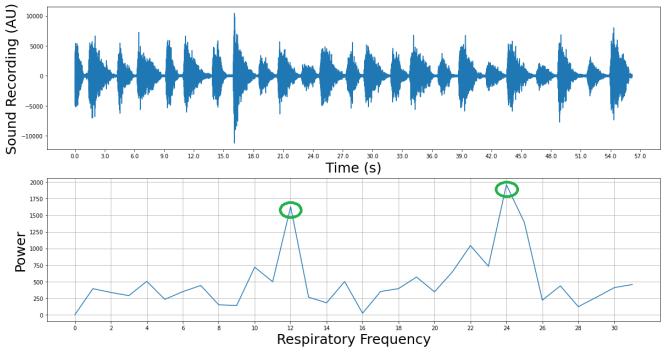


Fig. 1: Top: an audio clip of breathing which contains 12 breaths; bottom: a Fast Fourier Transform of the audio clip. The peak with the highest power was found at 24 BPM, and a corresponding prominent peak was found at 12 BPM. The two peaks are offset by a factor of 2, indicating that the respiratory rate is 12 BPM, with at peak at 24 BPM representing the number of inhales and exhales combined.

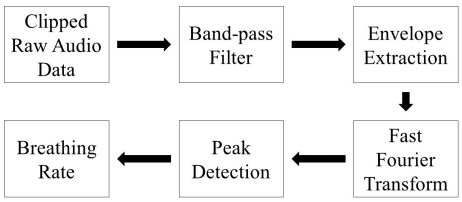


Fig. 2: A simple block diagram of the respiratory rate estimation algorithm.

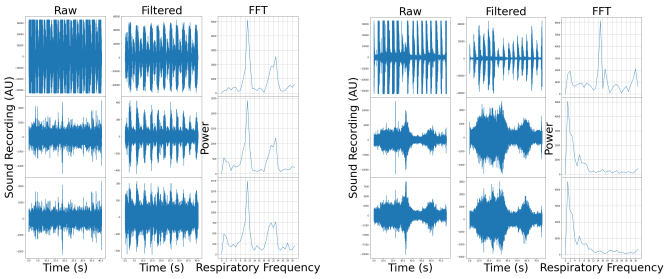
the PE to determine if there was a significant difference between the three microphone placements. If a significant difference was found, the SNR and the PE would then be compared in pairs between the three microphone locations. To compare the data quality between the different microphone placements a paired Wilcoxon signed-rank test was applied with a Bonferroni correction for multiple comparisons.

C. Results

The breathing clips included a mix of mouth and nose breathing. The respiratory rates contained within these clips ranged from 7 to 19 breaths per minute (BPM), with an average of 12.6 BPM (root mean squared deviation, RMSD = 3.3 BPM). The mouthguard-based microphone out performed the headgear-based microphone as well as the control microphone under both indoor and outdoor conditions, while the headgear-based microphone and the control performed similarly. Figure 3 demonstrates the difference in data quality between different microphones under both indoor and out door conditions.

A Friedman test on the indoor SNR dataset yielded a differences between the SNR of the three sensor placements under both indoor and outdoor conditions were found to be significant. Table 1 shows the details of the Friedman tests on SNR. Wilcoxon Signed-rank tests found that the SNR of the mouthguard placement is significantly higher than both the headgear placement as well as the control set, while the difference between the headgear placement and the control set is not significant, see table 2 for details of the tests.

Significant differences were found for respiratory rate between the sensor placements under both indoor and outdoor conditions according to Friedman tests (see table 3). The Wilcoxon Signed-rank



(a) Indoor (b) Outdoor

Fig. 3: These graphs show samples of time series data captured by microphones placed in different locations to visual demonstrate the difference in data quality. The top row shows mouthguard data, the mid row shows head gear data, and the bottom row shows external control data. It is visually clear that mouthguard data contain the least amount of noise and have the best signal strength.

Table 1: Signal-to-noise Ratio Friedman Test with N indicating the number of audio clips used.

	Indoor	Outdoor
N	40	40
Chi-Squared	43.35	33.8
Degrees of Freedom	2	2
P-value	3.86×10^{-10}	4.58×10^{-8}

Table 2: Signal-to-noise Ratio Wilcoxon Signed-rank Test

		Mouthguard VS Head Gear	Mouthguard VS External Control	Head Gear VS External Control
Indoor	Z-stat	5.00	5.00	0.44
	P-value	5.73×10^{-7}	5.73×10^{-7}	0.66
Outdoor	Z-stat	5.19	5.00	0.89
	P-value	2.12×10^{-7}	5.73×10^{-7}	0.375

Table 3: Respiratory Rate Algorithm Friedman Test with N indicating the number of audio clips used

	Indoor	Outdoor
N	40	40
Chi-Squared	47.41	55.96
Degrees of Freedom	2	2
P-value	5.07×10^{-11}	7.07×10^{-13}

Table 4: Respiratory Rate Algorithm Wilcoxon Signed-rank Test

		Mouthguard VS Head Gear	Mouthguard VS External Control	Head Gear VS External Control
Indoor	Z-stat	-4.78	-5.07	-0.08
	P-value	1.72×10^{-6}	3.97×10^{-7}	0.94
Outdoor	Z-stat	-5.32	-5.36	0.51
	P-value	1.06×10^{-7}	8.36×10^{-8}	0.61

tests suggest that the mouthguard sensor placement results in lower errors than both the headgear placement and the control set, while the headgear placement and the control set performed more alike (see table 4).

D. Discussion signal quality

The oral-cavity sensor placement captures higher quality data than external sensor placements. A simple frequency-domain-analysis algorithm without any optimization was applied for BPM estimation to minimize biases towards a particular microphone placement. The pre-processing steps are similar to those found in literature [3], [5], [6]. The resulting algorithm, when applied to the mouthguard dataset, had a relative error of 3.86% (RMSD = 9.30%) under indoor conditions, and a relative error of 4.58% (RMSD = 11.28%) under outdoor conditions. It was noted, qualitatively, that the estimations with higher errors resulted from audio data that were less uniform in either amplitude or breath duration. A further Wilcoxon Signed-rank test on the SNRs of indoor and outdoor mouthguard datasets resulted in a p-value = 0.54, indicating there is no significant difference in the data quality under indoor versus outdoor conditions.

III. AUTOMATED RESPIRATORY RATE EXTRACTION ALGORITHM

To further investigate the viability of an oral placement for monitoring respiratory rates, a machine-learning (ML) algorithm was trained to remove non-breathing sounds from the recorded audio. The audio cleaned by the ML algorithm will then be processed by the respiratory rate algorithm described above to give a respiratory rate estimate.

A. Data Acquisition

Two additional volunteers, one male and one female (aged 42 and 40), were recruited to obtain forty continuous breathing audio segments, around a minute each in length, with regular interruptions in between them in the form of speech and miscellaneous non-breathing respiratory sounds (such as coughing and throat-clearing). Data was collected under both indoor and outdoor conditions using the oral-based sensor location. This additional data was then used to as test set for assessing the performance of the proposed ML algorithm.

B. Data Processing

It has been shown that Convolutional Neural Network (CNN) can be used to classify respiratory sounds [7], and is easy to train given a limited data size. A CNN-based classifier with a moving window was designed to clean up recorded respiratory audio clips for further respiratory rate analysis. A window size of 10s with an 8s overlap was selected for the algorithm's sliding window. The 10s window was chosen based on the lowest expected breathing rate from literature, 6 breaths per minute [8]. The 8s overlap was determined by the expected length of common noises within audio data of breathing, which was estimated to be around 2s.

The training data obtained were cropped into 2410 10s audio clips labelled as "pure breathing", and 1798 10s audio clips labelled as "containing noise". The breathing clips included a mix of nose and mouth breathing, and the noise clips included noises such as talking, coughing, throat clearing, drinking, swallowing, taking off/putting on the mouthguard, and ambient sounds while holding breath. The 10s clips were filtered with an 4th-order zero-phase Butterworth 100Hz to 400Hz band-pass filter. Data were downsampled to 1600Hz before they were transformed into spectrograms via Welch's method. A 3-layer CNN binary classifier was then trained with a 70/30 training/validation split. With a batch size of 32 and an early-stopping callback, the training was terminated after 18 epochs.

When an unedited audio clip is put through the trained algorithm, a sliding window method is applied to the clip, with filtering and then

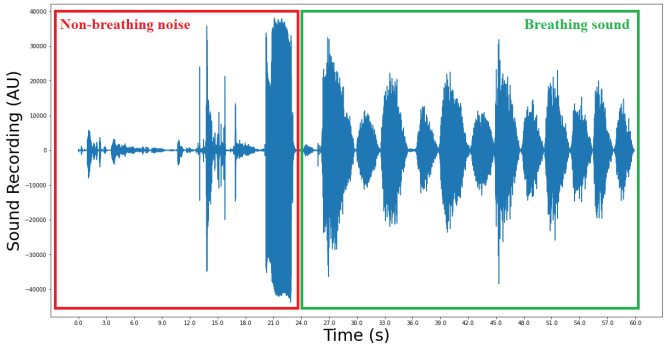


Fig. 4: In a real-world user scenario, the recorded audio would include non-breathing noises such as speech in addition to breathing signals. The CNN-based noise extraction algorithm can identify those non-breathing noises (highlighted in red) and breathing signals (highlighted in green), and crop the audio accordingly so that only breathing signals are fed into the respiratory rate estimation algorithm.

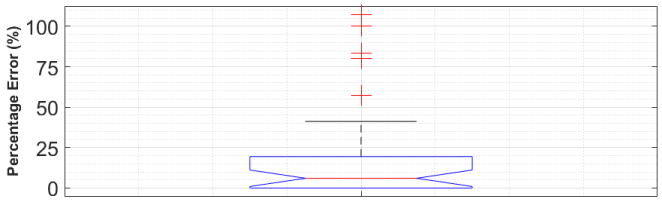


Fig. 5: Percentage error of respiratory rate (Rf) estimations on continuous breathing segments cropped by the CNN-based algorithm. A median of 6.07% relative error, with a 75th percentile of 19.44% relative error was found.

classification at each step. Once the sliding window moves through the entire clip, timestamps of continuous breathing periods within the audio clip are returned. These timestamps are then used with the FFT-based algorithm described previously to estimate the respiratory rate. Figure 4 depicts a visual representation of the output from the CNN algorithm.

C. Classifier

The trained CNN classifier had a training accuracy of 98.06%, and a validation accuracy of 95.65%, when applied to the test sets captured from two additional subjects that were unseen by the CNN classifier during training, the data clean-up algorithm was able to correctly extract 36 out of the 40 continuous breathing segments contained in the test data. The 36 extracted breathing audio segments were then put through the respiratory rate estimation algorithm, figure 5 shows the accuracy of the respiratory rate estimations.

IV. CONCLUSION

It was shown that, by placing an acoustic sensor in the oral cavity, one could obtain high-quality respiratory rate data. Furthermore, the higher data quality enabled a simple FFT-based algorithm to extract respiratory rates with a good accuracy. These results demonstrate the potential of a smart mouthguard for capturing respiratory rates in contact sports. Though, it was noted that when looking at the respiratory rate estimations individually, the FFT-based algorithm has a tendency to under-perform when the data is less uniform in either amplitude or breath duration, which will potentially impact the algorithm's ability to perform in contact sports settings. So although the data quality has been shown to be sufficient in this study, the method needs to be verified in a data sample that's more representative for the overall contact sports population. In addition, if this data processing pipeline were to be implemented in real-time, it would require up to a minute-long time delay, depending on the respiratory rate, so there's enough data to be parsed at once.

More sophisticated algorithms should also be explored. Machine learning approaches for time series audio analysis have been developed for speech-to-text applications in natural language processing (NLP), with some common examples being Recurrent Neural Network (RNN), Hidden Markov Model (HMM), and Transformer architecture [9]. Furthermore, studies have shown that RNNs can be used to detect breathing events for monitoring sleep disorders [10] or during speech [11]. Thus, it is possible to borrow techniques, such as RNN, from speech-to-text NLP to extract respiratory rate as well as identify respiratory noises, such as coughing, that might be of medical interest. Moving forward, a more comprehensive well-labelled breathing audio data set needs to be collected, in order to train a model that extracts respiratory rate and identifies respiratory noises, which will allow for respiratory rate monitoring in contact sports settings in real-time.

ACKNOWLEDGMENT

This project was funded by the EPSRC Impact Acceleration Grant EP/R511742/1 and partly funded by a Lab10X grant (reference OUI18284)

REFERENCES

[1] R. Cheng and J. H. Bergmann, "Impact and workload are dominating on-field data monitoring techniques to track health and well-being of team-sports athletes," *Physiological Measurement*, vol. 43, no. 3, p. 03TR01, 2022.

[2] A. Nicolò, S. M. Marcora, and M. Sacchetti, "Respiratory frequency is strongly associated with perceived exertion during time trials of different duration," *Journal of sports sciences*, vol. 34, no. 13, pp. 1199–1206, 2016.

[3] L. de Almeida e Bueno, M. T. Kwong, W. R. Milnthorpe, R. Cheng, and J. H. Bergmann, "Applying ubiquitous sensing to estimate perceived exertion based on cardiorespiratory features," *Sports Engineering*, vol. 24, no. 1, pp. 1–9, 2021.

[4] A. Nicolò, M. Montini, M. Girardi, F. Felici, I. Bazzucchi, and M. Sacchetti, "Respiratory frequency as a marker of physical effort during high-intensity interval training in soccer players," *International journal of sports physiology and performance*, vol. 15, no. 1, pp. 73–80, 2020.

[5] Y. Nam, B. A. Reyes, and K. H. Chon, "Estimation of respiratory rates using the built-in microphone of a smartphone or headset," *IEEE journal of biomedical and health informatics*, vol. 20, no. 6, pp. 1493–1501, 2015.

[6] A. Martin and J. Voix, "In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1256–1263, 2017.

[7] F. Barata, K. Kipfer, M. Weber, P. Tinschert, E. Fleisch, and T. Kowatsch, "Towards device-agnostic mobile cough detection with convolutional neural networks," in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2019, pp. 1–11.

[8] G. Benchetrit, "Breathing pattern in humans: diversity and individuality," *Respiration physiology*, vol. 122, no. 2-3, pp. 123–129, 2000.

[9] S. Karita, N. Chen, T. Hayashi, T. Hori, H. Inaguma, Z. Jiang, M. Someki, N. E. Y. Soplin, R. Yamamoto, X. Wang *et al.*, "A comparative study on transformer vs rnn in speech applications," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2019, pp. 449–456.

[10] E. Urtnasan, J.-U. Park, and K.-J. Lee, "Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal," *Neural computing and applications*, vol. 32, no. 9, pp. 4733–4742, 2020.

[11] V. S. Nallanthighal, A. Härmä, and H. Strik, "Speech breathing estimation using deep learning methods," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1140–1144.