CoBrS: Cough Breath Segmentation for the reduction of class-confounding characteristics in dataset curation

Alice E. Ashby
Department of Engineering & Informatics
University of Sussex
Brighton, United Kingdom
A.Ashby@sussex.ac.uk

Khuong An Nguyen
Department of Computer Science
Royal Holloway University of London
Surrey, United Kingdom
Khuong.Nguyen@rhul.ac.uk

***Abstract*—Cough segmentation using Machine Learning is known to be sensitive to the effects of class-confounding characteristics in the training data, significantly skewing predictions with the introduction of bias. Mechanisms by which bias may permeate a dataset include small sample sizes and noise in the samples. In this paper, we propose a novel audio segmentation algorithm as a means to solve these issues through automatic isolation and extraction of biological audio events. Our algorithm, CoBrS, is based on heuristics derived from physiological assumptions and is designed to accurately isolate all cough types, including the complex peal cough, and provides segmentation support for breaths, a previously undocumented modality in segmentation literature. CoBrS was validated on three public cough datasets with varying segmentation complexity (Coswara, COUGHVID, Virufy) against two state-of-the-art algorithms (COUGHVID and Virufy), achieving mean signal quality increases of 169.3%, 274.2%, and 39.8%, and sample size increases of 250% and 280% respectively. Our findings were also manually verified by two human raters who reported a 94% peal cough segmentation rate and that 88% of coughs in the moderate noise test subset are of high quality. Our algorithm is capable of effectively isolating cough and breath events of all types from samples with low to moderate noise, whilst improving signal quality and retaining high-frequency information that is often lost in the process.**

***Index Terms*—cough segmentation, audio signal processing, audio noise filtering, breath segmentation, bias mitigation**

I. Introduction

In recent years, diagnosis of respiratory diseases such as COVID-19 through the amalgamation of digital signal processing (DSP) and machine learning (ML) techniques on respiratory sounds has become a viable option for digital mass testing [1]–[3]. The classification performance of these ML algorithms relies in part on consistent, high-quality data, where predictors can effectively discriminate classes based on computed features. Class-confounding characteristics, introduced into the training data where some variables are uncontrolled, can significantly skew predictions and lower classification accuracy [4]. Clinical and crowdsourced bioacoustic datasets [5]–[7] often contain recordings of various durations containing many biological events that are unfortunately interspersed with noise, sometimes with noise overlapping the events, rendering it impossible to meaningfully extract latent features.

For example, Coppock et al. propose that bias can be subtly encoded in audio samples through the ambient acoustical environment in which the coughs were recorded [8]. Consider the classification of COVID-19 negative and positive coughs. Positive coughs are often recorded in medical settings, and this acoustical environment is consistent and embedded in the audio sample. An issue arises if a predictor learns this association in the training set and a positive cough is present in the test set outside of this learned environment; the predictor might misclassify it as a negative cough due to bias [8].

We hypothesise that an intuitive segmentation strategy can address these ubiquitous challenges affecting data quality in this emerging field, and propose such a strategy in this work.

A. Our Contributions

- We develop a novel audio segmentation algorithm based on heuristics derived from physiological assumptions, allowing us to accurately isolate a complex third cough type, the peal cough, which is often overlooked in other segmentation studies and consequently identified as individual coughs by hypersensitive algorithms.
- We incorporate the breath modality, which is often mixed within the cough signals, but whose waveforms present differently to coughs, with varied patterns, lower amplitudes, and unique frequencies. To our knowledge, this is the first segmentation algorithm for the breath modality and the first validation study on a breath dataset.
- We validate our segmentation algorithm on three public datasets with varying segmentation complexity against two state-of-the-art algorithms. Additionally, two human raters manually spot-checked all segmented subsets to ensure the quality of the proposed method.

The remainder of the paper is organised as follows. Section II explores related work in the field of cough audio segmentation and how our work expands upon these implementations. Section III disseminates the physiological information regarding cough and breath events and how our proposed algorithm is developed with this information in mind. Section IV outlines empirical experimentation on three public datasets,

to extensively validate our proposed algorithm. Section V explores the results of our validation and how we address the problems outlined in Section I. Lastly, Section VI provides a summary of our results and our ambitions for the future.

II. Related Work

ML-based audio segmentation is still in its infancy. Cohen-McFarlane et al. explored different methods for silence removal and segmentation optimisation for cough events, using indoor-recorded audio with low to moderate background noise [9]. They compared standard deviation (SD), short-term energy (STE), and zero-crossing rate (ZCR) to manual segmentation, and found that ZCR had the highest accuracy at 89% but failed under noisy conditions [9]. Their research was promising, and so we built upon this by further assessing the effects of segmentation on sample size and signal quality.

Orlandic et al. developed a cough segmentation algorithm as part of a quality estimator for the COUGHVID cough sounds database [6]. The algorithm is based on a digital hysteresis comparator on the signal power, firstly preprocessed through normalisation to the [-1, 1] range, downsampling to 12 kHz, and lowpass filtering ($f_{cutoff} = 6$ kHz) [6]. This large-scale study differs from us in how audio signals are pre-processed and segmented, explored in detail in Sections III-B & IV-A. Their algorithm is publicly available, and so we consider it state-of-the-art and validate our results against theirs.

Lastly, Amrulloh et al. proposed a complex automated cough extraction methodology for pediatric ambulatory recordings using deep learning (DL) [10]. A feature matrix including Mel-frequency cepstral coefficients (MFCCs), Formant frequency, ZCR, non-Gaussianity score (NGS), and Shannon entropy is fed into a trained artificial neural network (ANN) classifier to differentiate sub-blocks into cough and non-cough classes. In this paper, we offer an alternative without necessitating a need for DL computational overhead, supporting inclusion in existing workflows and usage in edge devices.

III. Methodology

We hypothesise that an intuitive segmentation strategy can address the ubiquitous challenges affecting data quality outlined in Section I, by modelling the physiological characteristics of these biological events and extracting them in their entirety from the main audio sample; thus, a new child sample is created without noise and encoded bias, therefore improving the signal quality. By only extracting clear, meaningful data from the audio sample and not the events overlapping with noise, the signal quality of the overall training set is improved; this is shown to correlate with better classification accuracy.

Consequently, segmentation increases the training set sample size. Audio samples with a poor quality ratio of biological events are no longer discarded completely; good-quality events are first extracted before discarding the low-quality remainder.

A. Cough and Breath Physiology

Understanding the physiology of cough and breath sounds is paramount when designing a segmentation algorithm that will not detect non-biological sounds or ambient noise. The physiological mechanics of a cough consist of a stereotypic inspiratory phase, then complete closure of the glottis, allowing compression of the thorax to increase subglottic pressure (compression phase). Then, the rapid opening of the glottis results in a peak supramaximal expiratory flow phase followed by high expiratory airflow (plateau phase) [11]. However, we govern our algorithm design by the observable phases, not the airflow phases. The inspiratory/compressive phases are near-silent, followed by peak airflow and plateau expiratory phases (termed phase 1: the explosive phase), with an intermediate phase (phase 2) and a voiced phase (phase 3) following. Coughs with three distinctive phases are characterised as 3-phase coughs, and coughs without a voiced phase are known as 2-phase coughs; two phases occur in approximately one-third of subjects [11]. We explore these varying phase configurations and the different inter-phase energy fluctuations in our previous work [3], [12]. Previous studies (including ours), however, neglect a complex third cough type known as a peal cough (see Figure 1, Graph A). Therefore, it is the focus of this study. A peal cough presents with a large inspiration followed by multiple expulsive events, each typically decreasing in expired volume. Unlike the average duration of 410 ms for a standard cough, peal coughs are much longer averaging 1.23 s [11].

The second focus of this study is breaths (see Figure 1, Graph B). The physiological mechanics of breath consist of an inspiratory phase, where the diaphragm and ribcage expand and contract allowing airflow and pressure balance, then exhalation to relax the diaphragm and expel air (post-inspiratory phase) [13]. We found that the duration of the inspiratory phase is approximately 1 to 1.5 s, with a very brief intermediary phase of 100 to 300 ms, then post-inspiratory 1.5 to 2 s. We also analysed the breathing patterns of subjects in the data; the majority were box breathing, where there is an equal duration of inhalations and exhalations [14]. In lesser numbers, we observed cyclic sighing, which emphasizes prolonged exhalations, and cyclic hyperventilation with retention, where there are longer inhalations and shorter exhalations [14]. Through manually listening to samples we found there were various combinations of nasal, oral, and nasal-oral breathing present in the data. We are the first to perform such analysis.

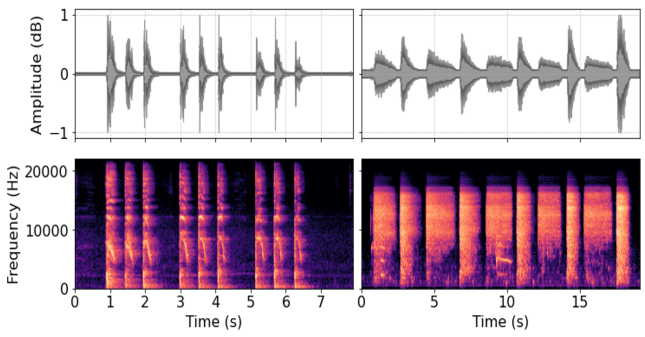

Fig. 1. Time-expanded waveform and spectral analysis of cough & breath demonstrate frequency spread up to 20 kHz, supporting existing research [15].

B. Our Proposed Algorithm

Our proposed algorithm focuses on the physiology of coughs (particularly the underrepresented peal coughs) and breaths. We propose assessing when a cough is present based on an energy threshold being exceeded, then retaining the information contained within until the cough is detected as having ended. In the case of overlapping detected events due to hypersensitivity of multiple energy fluctuations, we will merge them to form one detected event. Hypothetically, this allows for cleaner segmentation of variable-length peal coughs.

We begin with iterating through the input audio signal ($signal_n$ where the signal is an array of n amplitudes) by equally sized temporal slices, where each 50 ms slice is mapped to an envelope containing the following:

$$\bar{e_v} = signal_n \left[slice_{start} : slice_{end}\right]^2 \tag{1}$$

$$e_t = \frac{slice_{start}}{f_s} \tag{2}$$

Where $\bar{e_v}$ refers to the mean averaged $envelope_{value}$, e_t refers to $envelope_{time}$, and $f_s = 44100$. We select an $envelope_{size}$ of 50 ms to get an accurate segmentation of the peal cough. The intuition behind this is that there is a variable 10 to 30 ms of silence between peal epochs, and an $envelope_{size} < 0.05$ would consider this to be the end of the present cough currently iterated over, with the next envelope to contain a new cough. This would result in the clipping problem, where peal coughs are partially clipped, resulting in a loss of class-discriminative information. We confirmed this hypothesis through validation experiments on the data.

Next, we compute the noise floor N_{floor} of the full input audio signal in Equation 3 using a specified detection threshold $threshold_{peak} = 1.5$, selected by extensive validation. The N_{floor} measures the signal created from the sum of all the noise sources and unwanted signals. We consider the entire signal to compute this, as the recordings are generally consistent in background noise; we observed this through manual listening of audio samples, where we found either a lot of noise across the signal or not so much. It is defined as follows:

$$N_{floor} = \left(\frac{\sum_{i=1}^{N} signal_i}{N}\right)^2 \cdot threshold_{peak} \tag{3}$$

We then filter the envelopes above the N_{floor} to find the peaks of the audio signal and refine the event span (total duration of the audio event) by finding the onset and offset of the audio segment that contains the peak. This is achieved by using the detected peak $peak_{time} \cdot f_s$ as a starting point and stepping forward across the signal (via $envelope_{size} \cdot f_s$) until $\bar{e_v} < N_{floor}$, meaning the onset of the event has been located. The inverse is then performed by stepping backwards across the signal until $\bar{e_v} < N_{floor}$ to locate the offset.

We then filter only the events that conform to $onset_{peak} \geq onset_{valid}$ where $onset_{valid}$ is defined by the following:

$$signal_n \left[onset_{peak} - (time_{val} \cdot f_s)\right] + threshold_{val} \tag{4}$$

Where $threshold_{val} = 6$ and $time_{val} = 0.125$ (125 ms). This ensures that some moderate background noise with initially high energy peaks will not be segmented. Both $threshold_{val}$ and $time_{val}$ are selected based on validation.

Percussive peaks contiguous to the signal boundaries are filtered out. These peaks are observable in audio waveforms but not audible to the human ear; we consider them as noise and remove them as such. In addition, we define a minimum event duration based on physiological parameters, $length_{min} = 0.01$ (100 ms), where an accepted biological event is defined by Equation 5 and anything less is discarded.

$$\frac{\Delta onset}{f_s} \geq length_{min} \tag{5}$$

To handle peal coughs more effectively, we merge any overlapping events into a single event. This is achieved by sorting an array of candidate events by their onset, iterating through the array, and checking whether $onset_{current} \leq offset_{previous}$. If *false*, then the iteration continues; if *true*, then the current onset is omitted and the offset related to that onset joins the previous onset, combining the two events. This eliminates the hypersensitivity observed in most segmentation algorithms that do not take peal coughs into account. The resulting array of merged min-length conforming events (denoted as onsets and offsets) are then combined where $signal_n \left[onset : offset\right]$ is a segmented cough or breath.

IV. Experiments

In this section, a comprehensive evaluation is performed to investigate the performance of the proposed algorithm.

A. Data Acquisition and Pre-Processing

We acquire two public cough datasets and one public multimodal (cough and breath) dataset to demonstrate the validity of our proposed model for cough and breath segmentation; COUGHVID [6], Coswara [5], and Virufy [7].

TABLE I
SUMMARY OF THE THREE EXPERIMENT DATASETS.

Dataset	**Noise**	**Duration (s)**	Σ	f_s **(kHz)**
COUGHVID	Low	8.7	100	48
Coswara (C/B)	Moderate	5.8/16.8	100/65	Varies
Virufy	High	18.9	16	48

For consistency, we convert all WEBM, OGG, and MP3 files to 16-bit PCM WAV files. Next, we convert samples to mono-channel and normalise by scaling the amplitudes to the [-1, 1] range to account for intra-sample differences in volume. As we found that cough and breath sounds can have frequency components spread up to 20 kHz (see Figure 1), we resample the audio to 44.1 kHz, which would include frequency information up to 22.05 kHz according to the Nyquist-Shannon theorem without distortion from aliasing. Note that Coswara has mixed sample rates f_s of 48 kHz, 44.1 kHz, and 16 kHz, where 48 kHz is dominant and 16 kHz

Algorithm 1 Event detection and filtering algorithm.

Require: $signal_n$: audio signal, $envelope_{size}$: envelope size $threshold_{peak}$: detection threshold, f_s: sample rate, $time_{val}$: validation time, $length_{min}$: minimum segment length, $threshold_{val}$: validation threshold.
Ensure: $segmented\_signal_n$: segmented audio signal

$slices, envelopes, events \leftarrow$ List()
for $slice$ **in** Range(0, Length($signal_n$), ($envelope_{size} \cdot f_s$)) **do**
  $slices$.append(Tuple($slice$, ($slice + (envelope_{size} \cdot f_s)$)))
end for
for $slice$ **in** $slices$ **do**
  $envelope_{value} \leftarrow$Power(Mean($slice_n$), 2)
  $envelope_{time} \leftarrow (slice / f_s)$
  $envelopes$.append(Dict($envelope_{value}, envelope_{time}$))
end for
$N_{floor} \leftarrow$Power(Mean($signal_n$), 2) $\cdot$ $threshold_{peak}$
$detected_peaks \leftarrow$List($envelope$ **for** $envelope$ **in** $envelopes$ **do if** $envelopes\,[envelope_{value}] > N_{floor}$)
for $peak$ **in** $detected\_peaks$ **do**
  $current_i \leftarrow (peak\,[envelope_{time}] \cdot f_s)$
  $step \leftarrow (envelope_{size} \cdot f_s)$
  while $(current_i + step) <$ (Length($signal_n$) $- 1$) **do** ▷ Note: this *while* loop is repeated twice to get max_i and min_i
    $past_i \leftarrow current_i - step$
    $future_i \leftarrow current_i + step$
  end while
  $max_i \leftarrow$Length($signal_n$) $- 1$
  $events$.append(Dict($peak\,[envelope_{time}] \cdot f_s), onset \leftarrow min_i, offset \leftarrow max_i$)
end for
for $event$ **in** $events$ **do**
  if $(signal_n\,[event\,[peak]] \geq threshold_{val} + (signal_n\,[event\,[peak] - (time_{val} \cdot f_s)]$ **then**
    $filtered\_events$.append($event$)
  end if
end for
$filtered\_events$.sort($key \leftarrow$**lambda** i: $i\,[onset]$)
for $event_{cur}$ **in** $filtered\_events$ **do**
  $event_{prev} \leftarrow filtered\_events\,[-1]$
  if $event_{cur}\,[onset] \leq event_{prev}\,[offset]$ **then**
    $event_{prev}\,[offset] \leftarrow$ Max($event_{prev}\,[offset], event_{cur}\,[offset]$)
  else
    $merged\_events$.append($event_{cur}$)
  end if
end for
for $event$ **in** $merged\_events$ **do**
  if $((event\,[offset] - event\,[onset]\,/ f_s) \geq length_{min}$ **then**
    min_len_events.append($event$)
  end if
end for
for $event$ **in** min_len_events **do**
  return $signal_n\,[event\,[onset] : event\,[offset]]$
end for

are outliers. Due to the low prevalence of 16 kHz samples, these are discarded to prevent distortion from interpolation to a higher f_s. Samples with $f_s = 44100$ we normalise only.

B. Evaluation Metrics

We compute a measure of the strength of the desired signal relative to an undesired signal, known as the signal-to-noise ratio (SNR). The SNR can be defined as follows:

$$SNR_{dB} = 20 \cdot \log_{10} \left(\frac{A_{signal}}{A_{noise}} \right)^2 \tag{6}$$

Where A_{signal} are the amplitudes of the signal samples corresponding to biological events of interest (coughs and breaths), and A_{noise} are the amplitudes of non-biological events of no interest, presumed to be ambient background noise. We compute the SNR where A_{signal} is the absolute mean $|\mu|$ of the signal and A_{noise} is the standard deviation σ. We use this metric to validate our hypothesis that segmented bioacoustic events will contain less noise than their parent samples from which they are isolated, as higher SNR scores are an intuitive indicator of signal quality improvement [16].

The Virufy segmented subset and all child samples where their parent contains a peal cough in the COUGHVID and Coswara subsets are manually spot-checked by two human raters to assess the segmentation quality under noisy conditions. The rating scale is as follows: 'high quality' for audio containing only a clear and complete cough, 'low quality' for audio containing either a partial cough or a full cough with noise, and 'discard' for audio containing no discernible cough.

C. Empirical Results

This section outlines the key results of our experiments on the COUGHVID, Coswara cough (Coswara-C), and Coswara breath (Coswara-B) subsets against a state-of-the-art (SOTA) algorithm COUGHVID (see Section II). For a fair comparison, we do not test the COUGHVID algorithm on Coswara-B, as it is clear that this algorithm was designed with only the cough modality in mind [6]. SNR scores are computed for the COUGHVID algorithm with the same method as CoBrS. In the following text, we denote mean signal quality as $\bar{s}_{snr}$, mean sample duration as $\bar{s}_{len}$, and sample size as s_{size}.

For CoBrS on the COUGHVID subset, $\bar{s}_{snr}$ increased by 775.8% and s_{size} by 280%. In comparison to the COUGHVID algorithm, CoBrS $\bar{s}_{snr}$ and s_{size} were 274.2% and 35.7% higher, respectively. We compute SNR scores for both positive and negative samples to analyse class-wise differences in $\bar{s}_{snr}$. We found that with CoBrS, class-wise differences in $\bar{s}_{snr}$ decreased from 98.5% to 3.3%, but only 32.4% for COUGHVID. We do the same for $\bar{s}_{len}$, and found that class-wise differences for CoBrS were reduced from 6.7% to 4.5%, whereas for COUGHVID, it increased to 15.9%.

For CoBrS on the Coswara-C subset, $\bar{s}_{snr}$ increased by 497.6% and s_{size} by 250%. In comparison to the COUGHVID algorithm, CoBrS $\bar{s}_{snr}$ and s_{size} were 169.3% and 45.8% higher, respectively. For CoBrS on the Coswara-B subset, $\bar{s}_{snr}$ increased by 672.7% and s_{size} by 213.9%. We found that class-wise differences in $\bar{s}_{snr}$ decreased from 70.8% to 14.3%.

The Virufy dataset includes an unprocessed subset and a segmented subset produced by their in-house SOTA segmentation algorithm. Their code is unreleased and we were unable to reimplement the algorithm; therefore, we segment the unprocessed subset with CoBrS and compare our results to the Virufy segmented subset. CoBrS achieved a $\bar{s}_{snr}$ 143% higher than the original samples and 39.8% higher than Virufy.

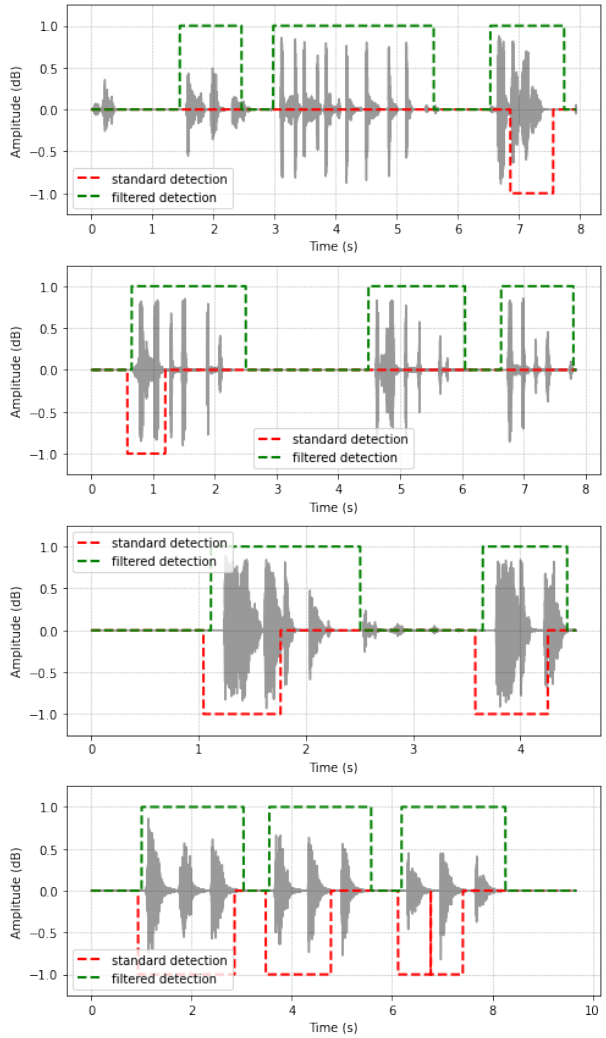

Fig. 2. Filtered onset detection (**CoBrS**) vs standard onset detection (**COUGHVID**) of peal coughs. **(a)** peal cough with 6 additional expulsions after the initial one. **(b)** 3 peal coughs with varying expulsions. **(c)** peal cough and 3-phase cough. **(d)** 3 standard peal coughs.

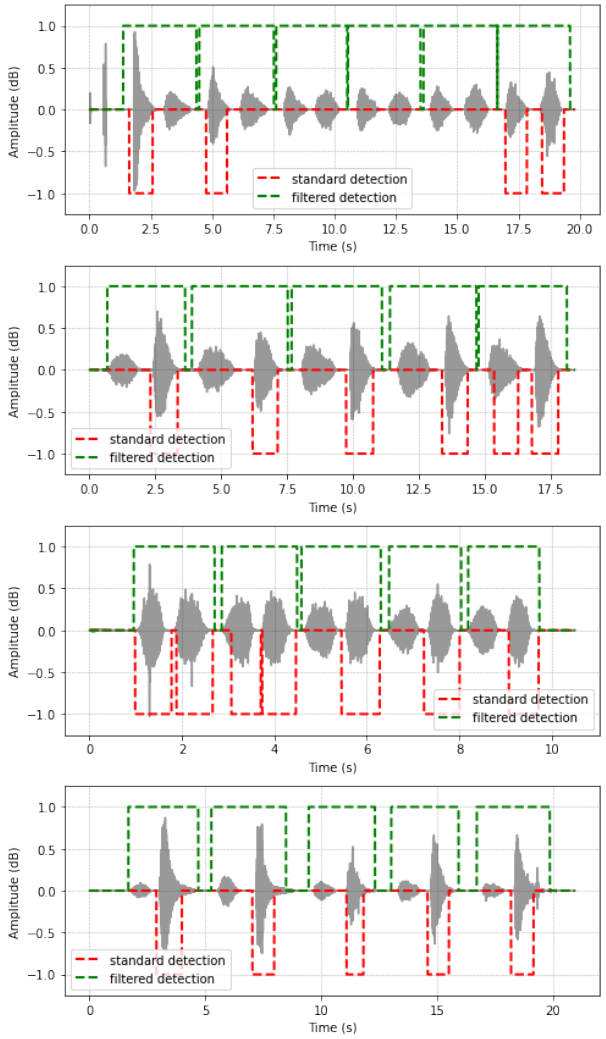

Fig. 3. Filtered onset detection (**CoBrS**) vs standard onset detection (**COUGHVID**) of breaths. **(a)** nasal, variable pitch, box breathing. **(b)** nasal, louder exhalation, cyclic hyperventilation. **(c)** nasal, variable pitch, box breathing. **(d)** oral, louder exhalation, cyclic sighing.

We also observed class-wise differences in $\bar{s}_{snr}$ were 124.9% for Virufy compared to 81% for CoBrS. In addition, the subset of Virufy segmented coughs generated by CoBrS ($n = 199$) was manually spot-checked by two human raters (see Section IV-B). They found that 88% of samples contained high-quality coughs despite the Virufy dataset containing the most noise.

V. Results Discussion

Overall, the proposed algorithm demonstrated promising results in segmentation capability in scenarios with variable ambient background noise. For example, Coswara had a lower innate signal quality than COUGHVID by 14.6%, and Virufy had 43.7% less signal quality than Coswara. Despite the incremental complexity, CoBrS achieved significant to moderate $\bar{s}_{snr}$ improvement across the board, outclassing current SOTA methods COUGHVID and Virufy by a wide margin. The following figures help to illustrate this improvement.

In Figure 2, we can discern the accuracy of our segmentation method visually for a diverse array of cough types, comparing our performance to SOTA. Of interest is the repeated neglect of peal coughs to be isolated by SOTA (see Graphs A & B) or the commonality of partial isolation (see Graphs C & D), a problem known as cough hypersensitivity or clipping.

In Figure 3, similar effectiveness is shown for various breath types; often SOTA are unable to identify shallow breaths interposed between the voiced phases, whereas CoBrS can isolate both phases together in the same sample, as desired.

In Figure 4, we showcase an upward trend where a higher proportion of SNR scores move away from 0, indicating better signal quality over all datasets, as noise is reduced.

In Figure 5, class-wise differences in $\bar{s}_{len}$ were reduced in the segmented dataset (6.7% to 4.5%), aligning the distributions more closely. Contrarily, SOTA increased by 15.9%.

Two human raters performed manual spot-checking on the

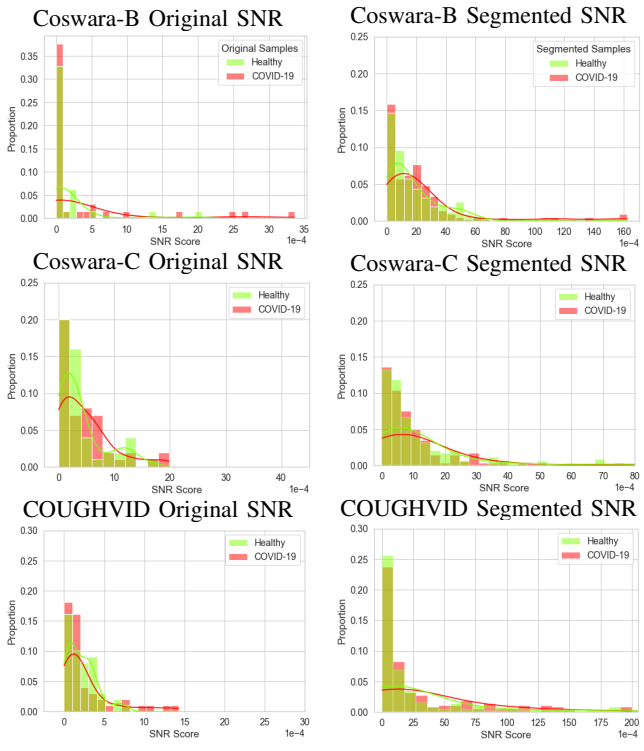

Fig. 4. SNR scores across the three datasets before and after segmentation.

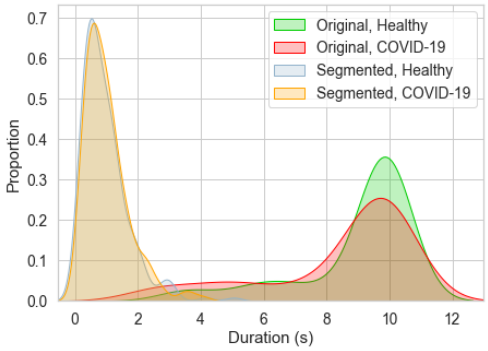

Fig. 5. Class-wise duration distributions before and after segmentation.

COUGHVID and Coswara subsets. In total, 180 segmented peal coughs ($n = 60$ parent samples, ≈ 3 coughs per sample) were analysed and a 94% peal cough segmentation rate was observed, based on the criteria outlined in Section IV-B.

These results support our previous research [3], [12], where we proposed a cough segmentation method based on cough clustering, which we found to address hypersensitivity to cough-phase energy fluctuations. We proved that segmenting a training set could improve classification of the minority class by up to 20% without detriment to the majority class [3].

VI. Conclusions and Future Work

In this paper, a novel segmentation algorithm based on heuristics derived from physiological assumptions is proposed to isolate a wide variety of biological acoustic events pertaining to two respective modalities, cough and breath. Our empirical, robust human-verified findings, show that CoBrS eclipses the respective performances of SOTA methods and is effective at segmentation in low to moderate noise conditions, simultaneously improving signal quality and retaining high-frequency information (up to 22.05 kHz) that is often lost in the segmentation pre-processing stages of other methods.

In the future, we plan to validate the robustness of the algorithm in more diverse settings, such as ambulatory recordings in hospital wards and on differing pathologies of bioacoustic samples, i.e. asthma, bronchitis, pneumonia, pertussis, and influenza. The effects of segmentation methods on the class imbalance problem should also be explored in detail. Lastly, we hypothesise that other bioacoustic events of interest, such as snoring in polysomnography, could benefit from similar segmentation strategies as we have proposed in this paper.

References

[1] Jing Han, Tong Xia, Dimitris Spathis, Chloë Brown, Ting Dang, et al., "Sounds of covid-19: exploring realistic performance of audio-based digital testing," *npj Digital Medicine*, vol. 5, pp. 16, 2022.

[2] Alexander Ponomarchuk, Ilya Burenko, Leonid Zhukov, et al., "Project achoo: a practical model and application for covid-19 detection from recordings of breath, voice, and cough," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 2, pp. 175–187, 2022.

[3] Alice Ashby, Julia Meister, Khuong An Nguyen, Zhiyuan Luo, and Werner Gentzke, "Cough-based covid-19 detection with audio quality clustering and confidence measure based learning," in *Conformal and Probabilistic Prediction with Applications*. PMLR, 2022, pp. 129–148.

[4] Shivani Gupta and Atul Gupta, "Dealing with noise problem in machine learning data-sets: A systematic review," *Procedia Computer Science*, vol. 161, pp. 466–474, 2019.

[5] Debarpan Bhattacharya, Neeraj Kumar Sharma, Debottam Dutta, Srikanth Raj Chetupalli, Sriram Ganapathy, et al., "Coswara: A respiratory sounds and symptoms dataset for remote screening of sars-cov-2 infection," *Scientific Data*, vol. 10, pp. 397, 2023.

[6] Lara Orlandic, Tomas Teijeiro, and David Atienza, "The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, pp. 156, 2021.

[7] Tara Wilson and Amil Khanzada, "Virufy covid-19 open cough dataset," https://github.com/virufy/virufy-data, 2020.

[8] Harry Coppock, Lyn Jones, Ivan Kiskin, and Björn Schuller, "Covid-19 detection from audio: seven grains of salt," *The Lancet Digital Health*, vol. 3, no. 9, pp. e537–e538, 2021.

[9] Madison Cohen-McFarlane, Rafik Goubran, and Frank Knoefel, "Comparison of silence removal methods for the identification of audio cough events," in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2019, pp. 1263–1268.

[10] Yusuf Amrulloh, Udantha Abeyratne, Vinayak Swarnkar, Rina Triasih, and Amalia Setyati, "Automatic cough segmentation from non-contact sound recordings in pediatric wards," *Biomedical Signal Processing and Control*, vol. 21, pp. 126–136, 2015.

[11] Kai Lee, Paul Davenport, Jaclyn Smith, Richard Irwin, Lorcan McGarvey, Stuart Mazzone, Surinder Birring, et al., "Global physiology and pathophysiology of cough," *Chest*, vol. 159, pp. 282–293, 2021.

[12] Alice Ashby, Julia Meister, Goran Soldar, and Khuong An Nguyen, "A novel cough audio segmentation framework for covid-19 detection," in *Proceedings of the Symposium on Open Data and Knowledge for a Post-Pandemic Era*. Electronic Workshops in Computing, 2022, pp. 1–8.

[13] Jan-Marino Ramirez, Tatiana Anderson, and Alfredo Garcia, "The ins and outs of breathing," *eLife*, vol. 3, pp. e03375, 2014.

[14] Melis Yilmaz Balban, Eric Neri, Manuela Kogon, and Andrew Huberman, "Brief structured respiration practices enhance mood and reduce physiological arousal," *Cell Reports Medicine*, vol. 4, no. 1, 2023.

[15] Akira Murata, Yasuyuki Taniguchi, Yasushi Hashimoto, and Shoji Kudoh, "Discrimination of productive and non-productive cough by sound analysis," *Internal Medicine*, vol. 37, no. 9, pp. 732–735, 1998.

[16] Sayanton Dibbo, Yugyeong Kim, and Sudip Vhaduri, "Effect of noise on generic cough models," in *International Conference on Wearable and Implantable Body Sensor Networks*. IEEE, 2021, pp. 1–4.