2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Audio-based cough counting using independent subspace analysis

Paul Leamy[1], Ted Burke[1], Dan Barry[2] and David Dorran[1]

***Abstract*— In this paper, an algorithm designed to detect characteristic cough events in audio recordings is presented, significantly reducing the time required for manual counting. Using time-frequency representations and independent subspace analysis (ISA), sound events that exhibit characteristics of coughs are automatically detected, producing a summary of the events detected without the need for a pre-trained model. Using a dataset created from publicly available audio recordings, this algorithm has been tested on a variety of synthesized audio scenarios representative of those likely to be encountered by subjects undergoing an ambulatory cough recording, achieving a true positive rate of 76% with an average of 2.85 false positives per minute.**

I. Introduction

The information contained within the sounds produced by the human vocal tract, such as speech, moaning, sighing, and coughing, present an opportunity to facilitate remote and non-contact monitoring of an individual's health [1]. In relation to physical health, coughing is a common symptom for which patients seek medical advice [2], especially in the acute and chronic categories [3], where a persistent cough can severely impair an individual's quality of life. In determining the frequency and severity of a person's cough, a clinician can make a suitable diagnosis relating to a person's cough, and this is the objective of a cough detection/monitoring system [4]. Having an objective measure of a cough is useful when tracking the progression of an illness [5] especially in the process of early detection and monitoring [6], highlighting the importance of cough monitoring systems to aid with tracking the progress of a disease.

Presented here is an algorithm designed to detect the presence of candidate cough sounds in audio recordings. The area of cough detection has received notable attention in recent years. Semi-automatic approaches using hand-crafted features as inputs to probabilistic neural networks (PNN) [7] and statistical models [8]–[11] achieved satisfactory results but rely on input from operators to successfully detect cough sounds. More recent algorithms have used eigenvalue decomposition with random forest classifiers [12], Hu moments with k-nearest neighbours (*k*-NN) [13], spectral features with time-delay neural networks (TDNN) [14], convolutional neural networks (CNN) and recurrent neural networks (RNN) [15], and gammatone filterbank cepstral coefficients (GFCC) with support vector machines (SVM) [16].

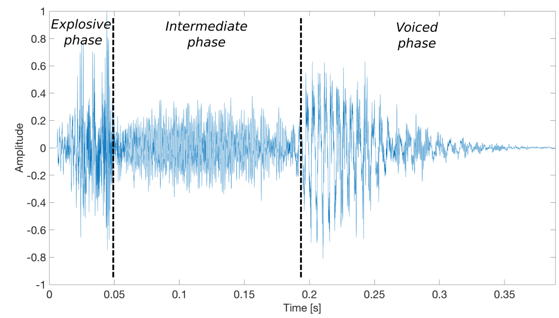

Fig. 1. Phases of a cough sound showing explosive phase, intermediate phase, and decaying voiced phase.

The nature of cough sounds, as described in [17], share characteristics with the signals produced by percussive and harmonic musical instruments. Cough sounds typically begin with a quick onset of wideband noise, stretching across a spectrum up to 20 kHz, followed by a short stationary period dependent upon the cause of the cough. An example of the phases associated with a cough can be seen in Figure 1. The onset of coughs share similar properties to the percussive characteristics of drums, and in [18] the transient nature of drum sounds were exploited using *independent subspace analysis* (ISA) [19] to automatically transcribe drum tracks from audio recordings. Due to the nature of ambulatory cough recordings, cough sounds produced by a person are likely to be more prominent within these recordings. The repeated occurrences of a person's cough and proximity to a microphone means that the cough sound is likely to be one of the sources contributing most variance to the recorded signal. Given the variance based nature of ISA, it is expected to produce candidate time-activation functions that coincide with cough events. An example of this algorithm's output is illustrated in Figure 2, which highlights the time-activation functions and corresponding frequency spectra for a short drum loop containing three sounds (kick drum, snare, hi-hat). The approach presented here builds on and refines the ISA approach for the purpose of cough event detection.

II. Methodology

Many existing approaches to cough detection are trained in a manner that aims to achieve generalisation such that algorithms can correctly identify cough sounds in a wide variety of recording scenarios. While generalisation is desirable in detection algorithms, it may not be required to achieve

Identify applicable funding agency here. If none, delete this.

[1]Paul Leamy, Ted Burke, and David Dorran are with the Biomedical Research Group at the Technological University of Dublin, Grangegorman lower, Dublin 7, Ireland {paul.leamy,ted.burke,david.dorran}@tudublin.ie

[2]Dan Barry is with QxLab at University College Dublin, Dublin, Ireland dan.barry@ucd.ie

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

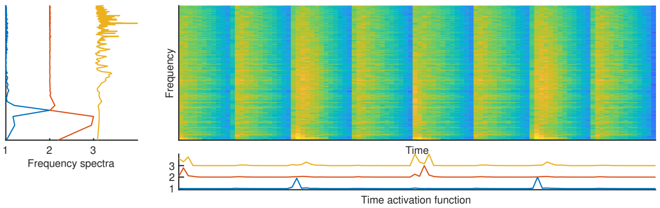

Fig. 2. Illustration of ISA being applied to a short drum loop, highlighting the significant time-activation functions and corresponding frequency spectra of each drum.

satisfactory results when detecting the presence of repeating events in more constrained scenarios. In ambulatory cough recordings, the coughs present are likely to come from a single person, which presents an opportunity, namely the characteristics of the coughs to be detected in an audio recording are likely to share similar time-frequency domain features.

The algorithm proposed in this paper is designed to summarise an ambulatory recording, producing a number of short audio clips, each of which is likely to contain a cough event. Using singular value decomposition (SVD) to analyse the spectrogram of an audio signal, a set of time-activation functions is produced which then undergoes independent component analysis (ICA). Together, SVD followed by ICA is known as independent subspace analysis. Significant peaks in these time-activation functions are then used as markers for candidate cough events since these peaks correspond to the presence of high variance events in the audio recording. Using a suitable threshold, candidate events can be marked and extracted from the input audio signal and presented to the clinician for further analysis and verification. The algorithm overview is illustrated in Figure 3, and an implementation of the algorithm is available at [20].

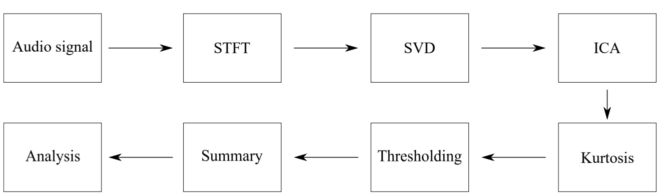

Fig. 3. Block diagram illustrating the process of the proposed algorithm. Further detail of each block in this process is presented in Section II of this paper.

A. Short-time Fourier transform

The first task of this algorithm is to compute the complex short-time Fourier transform (STFT) X of the input data x. This is computed using (1).

$$X(k,m) = \sum_{n=0}^{N-1} w(n)x(n+mH)e^{-j2\pi nk/N} \tag{1}$$

where k is the discrete-frequency index, m is the hop number for the analysis window, N is the frame size (2048 samples), and H is the hop size (512 samples). A Hanning window is used for the window function w, and the sampling frequency is $44.1\,\text{kHz}$. $X(k,m)$ is evaluated for $k = 0, ..., N/2$ and the magnitude of these components are retained.

B. Independent subspace analysis

SVD first decomposes the input matrix X into three matrices U, V, and S, to identify subspaces in order of the variance they contribute towards the original data [21]. In the context of the magnitude spectrum X, where columns contain the magnitudes of frequency content for a single analysis window, U are the frequency basis spectra and V are the corresponding time-activation functions. The outer product of each spectra and time-activation pair produces a subspace of the original input data, and the input data can be reconstructed using

$$X = USV^T \tag{2}$$

For this algorithm, the columns of V will act as time-activation functions for cough events in the input data X, and only the first nine singular values are computed (determined experimentally to be optimal).

Figure 4 illustrates the input and output of SVD on an example magnitude spectrum. These pairs of frequency-basis spectra and time-activation functions (U_1, U_2, V_1, V_1) both contribute towards reconstructing elements of both "sources" in the magnitude spectrogram. Ideally a single time-activation function would coincide with each single type of event has occurred, but in practice this is not the case. This can be overcome by applying ICA to the time-activation function to transform the decorrelated time-activations into statistically independent activation functions [22], [23]. This is the principle underlying ISA. In Figure 4, the plots on the right show the output of the ICA stage. The frequency-basis spectrum and time-activation function pair's contributions towards reconstructing both sources in the input magnitude spectrum has decreased significantly.

C. Time-activation selection and thresholding

Recall that in Section II-B the first nine singular values were computed. In determining which time-activation functions to retain, the kurtosis function $k(v)$ is used,

$$k(v) = \frac{\sum_{m=0}^{M-1}(V_{m,v} - \bar{V_v})^4/M}{\sigma_{V_v}^4} \tag{3}$$

where $V_{m,v}$ is the m^{th} sample in v^{th} column of V, $\bar{V_v}$ and σ_{V_v} are mean and standard deviation of v^{th} column of V, respectively. When normalised, kurtosis is 0 for a Gaussian distribution, positive for a "peakier" or leptokurtic distribution, and negative for a "flatter" or mesokurtic distribution. Sparse events are likely to occur sporadically in the time-activation functions, meaning a majority of near-zero values, resulting in a leptokurtic distribution. The time-activation functions with the three highest kurtosis measurements are retained and rectified to produce three candidate time-activation functions c_1, c_2, and c_3, for further analysis.

Peaks in each time-activation function correspond to moments in the input signal where the contribution of the corresponding frequency-basis spectrum is greater. A threshold

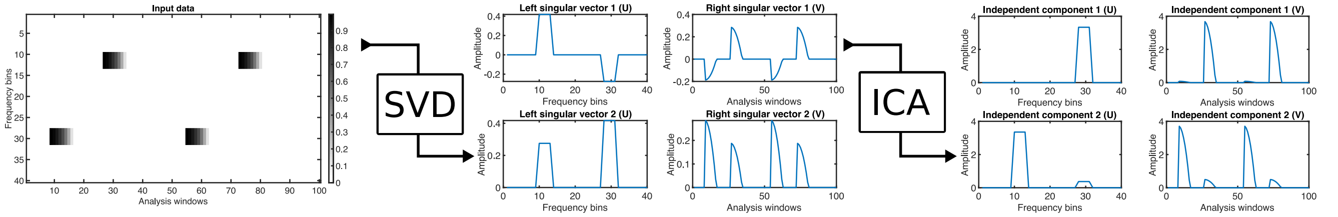

Fig. 4. Illustration of the ISA algorithm on a sample magnitude spectrum. Input data (left) undergoes SVD to produce the frequency basis spectra U and time-activation functions V. Applying ICA to the frequency-basis spectra and time-activation functions results in improved separation of the events, as can be seen in the four plots to right.

$\tau = a.\sigma_c$ is applied to the candidate time-activation functions to identify peaks, where σ_c is the standard deviation of c, and a is a constant, with $4 < a < 8$.

Retained peaks from the candidate time-activation function are used to generate a summary of the input signal, where $1\,\mathrm{s}$ windows around each peak are extracted from the input signal and concatenated to produce the short summary of the input signal.

D. Dataset and evaluation

The dataset used in evaluating the proposed algorithm was constructed from a number of publicly available cough, non-cough, and background sources collected from YouTube videos [24] and the DCASE 2016 Challenge [25]. Each test signal comprises 20 coughs produced by one individual and alternative foreground sounds (door knocks, table banging, speech, laughing, etc.). In total, 10 test signals were created with a duration of 10 minutes each. Instructions for reproducing these test signals, including annotations and URL links, can be accessed at [20].

The evaluation framework used here is adapted from [26], a framework for polyphonic sound-event detection which overcomes the limitations of collar-based event decisions and labelling subjectivity by annotators. *True positive* (TP) or *false positive* (FP) decisions rely on the degree of overlap with annotated events.

The overlap t_o between an annotated event and detected event (see Figure 5) is measured and expressed as a ratio of a $500\,\mathrm{ms}$ window, based on the average duration of coughs presented in [27]. When this ratio exceeds the detection tolerance criteria ρ_{DTC}, the detected event is labelled a TP, otherwise FP is declared,

$$D = \begin{cases} TP, & \text{if } \frac{t_o}{500ms} > \rho_{DTC} \\ FP, & \text{otherwise} \end{cases} , 0 < \rho_{DTC} \leq 1 \qquad (4)$$

where D is a single detection and $\rho_{DTC} = 0.3$. Undetected annotated events are classed as *false negatives* (FN).

The performance of this algorithm is quantified using the true positive ratio (r_{TP}) and the false positive rate (R_{FP}) [26],

$$r_{TP} = \frac{N_{TP}}{P} \qquad R_{FP} = \frac{N_{FP}}{T_{dur}} \qquad (5)$$

where N_{TP} and N_{FP} are the number of true and false positives, P is the number of annotated cough events, and T_{dur} is the duration of the signal being analysed.

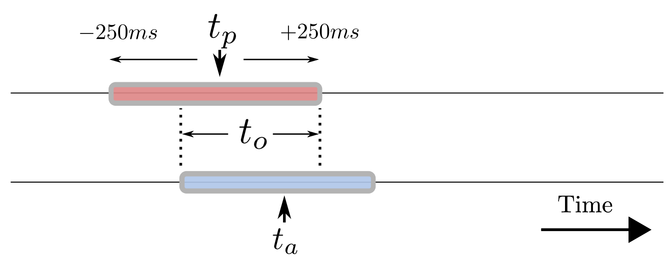

Fig. 5. Overlap between the time of a detected event t_p and time of an annotated event t_a is used to determine if a detected event is classified as a TP or FP. The overlap time t_o is expressed as a ratio of a $500\,\mathrm{ms}$ window. When this ratio exceeds the detection tolerance criteria ρ_{DTC} the detected event is marked as a TP.

III. RESULTS

The mean values for the true positive ratio, false positive rate, and summarised duration across all test signals are presented in Table II. These results were produced for each time-activation function. Activation function c_1 achieved the best performance in each metric across all mean values, suggesting this is the appropriate activation function to use. The increase in false positive rate with c_2 and c_3 suggests that these activation functions encapsulate features common to both cough and non-cough events after SVD.

Table III highlights a best-case scenario which includes the metrics for each test signal and the associated activation function that achieved this result. For the majority of results, c_1 produced the best case which aligns well with the results from Table II. For signal #3 an #5, the activation functions that achieved the most desirable results are c_3 and c_2, respectively. Assuming that the appropriate time-activation functions are used, an improvement in the mean true positive ratio and false positive rate is observed.

In Table I the performance of the proposed approach using ISA is presented and compared against several existing approaches in the field of *cough detection*, using reported values of the same metrics from (5). In some publications, these metrics were not reported by the original authors. Entries where this is the case have been labelled with **NR** for "not reported". Comments have also been provided on other existing approaches to highlight important aspects such as the requirement of training data and whether metrics are based on event level or frame level analysis, for example.

TABLE I

CROSS-COMPARISON OF COUGH DETECTION ALGORITHMS ON THE METRICS USED TO EVALUATE THE PROPOSED APPROACH. **NR** USED WHERE RESULTS WAS NOT REPORTED BY THE AUTHORS, NAMELY THE TRUE POSITIVE RATIO r_{TP} AND FALSE POSITIVE RATE R_{FP}.

Algorithm	Training Required?	r_{TP} (%)	R_{FP} (min^{-1})	Comment
Independent subspace analysis (our approach)	No	76.00	2.85	Metrics are based on cough events and training a model before testing is not required.
Hull automatic cough counter (HACC) [7]	Yes	80.00	**NR**	Semi-automatic approach to cough counting.
HMM (Leicester Cough Monitor) [8] HMM (Leicester Cough Monitor) [10]	Yes	71.00 91.00	0.22 0.04	Average results presented, semi-automatic algorithm that requires user input.
Spectral energy thresholds [11]	No	**NR**	**NR**	Results presented highlight the reduction in data to be manually analysed.
Eigenvalue decomposition and random forest classifier [12]	Yes	92.00	0.28	Cough model generated using training data and PCA. Average results reported.
TDNN and spectral features (event level) [14] TDNN and spectral features (frame level)	Yes	92.80 83.70	**NR** **NR**	Based on a frame-by-frame and event level. Definition of true negative is not stated.
Hu moments - synthetic recording and k-NN [13] with synthetic recording and SVM with ambulatory table-top recording with ambulatory pocket recording	Yes	82.49 62.83 87.43 81.83	**NR** **NR** **NR** **NR**	Statistics are calculated on a frame-by-frame basis.
STFT and CNN [15] STFT and RNN	Yes	86.88 87.70	**NR** **NR**	Reported results represent the best performance under the best-case scenario.
GFCC and spectral sub-band features (synthetic data) [16] with weighted voting (real data) with weighted voting and first derivatives (real data)	Yes	81.38 76.43 78.60	**NR** **NR** **NR**	Analysis carried out on sub bands using GFCC and SVM classifier. Average results calculated on a frame leve at SNR of 0 dB.

TABLE II

SUMMARY OF RESULTS FROM ALL CANDIDATE TIME-ACTIVATION FUNCTIONS (c_1, c_2, AND c_3) SHOWING THE MEAN RESULTS ACROSS ALL TEST SIGNALS.

c	$r_{TP}(\%)$	$R_{FP}(min^{-1})$	$T(min)$
1	76.00	2.85	0.73
2	67.00	4.17	0.92
3	39.00	6.26	1.17

TABLE III

A BEST-CASE SCENARIO REPRESENTATION OF THE RESULTS, WHERE c IS THE c^{th} ACTIVATION FUNCTION, r_{TP} IS THE TRUE POSITIVE RATIO, R_{FP} THE FALSE POSITIVE RATE, AND T IS THE SUMMARISED SIGNAL DURATION.

#	c	$r_{TP}(\%)$	$R_{FP}(min^{-1})$	$T(min)$
1	1	70	4.5	0.98
2	1	90	3.4	0.87
3	3	95	1.4	0.55
4	1	100	4.00	1
5	2	100	0.6	0.43
6	1	90	2.2	0.67
7	1	95	0.5	0.4
8	1	90	6.7	1.42
9	1	95	3.6	0.92
10	1	100	1.2	0.53
Mean		**92.50**	**2.81**	**0.77**

IV. DISCUSSION

From the results presented here, ISA has shown to be a reliable method for detecting the presence of cough sounds in audio recordings. A summary of the results from all candidate time-activation functions (c_1, c_2, and c_3) showing the mean results across all test signals is provided in Table II.

The first time-activation function, c_1, returned the best overall results with a true positive ratio of 76 %, false positive rate of 2.85 per min. In the case of the true positive ratio, the values obtained by the proposed algorithm outperform those reported in [8] with 71 % and in [13] of 62.83 % with synthetic recordings and a SVM classifier. In the analysis of real data in [16], the results are comparable where values of 76.43 % and 78.60 % were obtained (see Table I). Performance reduced slightly in the remaining candidate time-activation functions (c_2 and c_3), returning lower true positive ratio and higher false positive rates as less cough events were correctly identified and more false detections were made.

A best-case scenario is presented Table III. It can be seen that the majority of cough sounds are accurately detected within a single time-activation function, c_1 in this case, which supports the use of kurtosis as a suitable statistical measure for determining which activation function to use to achieve the most desirable results. On average, kurtosis produces the best activation function to use as shown in Table II. The mean results can be improved by selecting the time-activation function containing the coughs manually. An automatic method of determining which activation functions could be used to complement this algorithm and improve overall results, as shown by the mean calculations in Table III. This would improve the proposed algorithm's performance by accurately detecting cough events to $92.50\,\%$, outperforming several of the best performing approaches presented in Table I.

The false positive rate of the proposed algorithm could not be compared directly with a large number of the algorithms presented in Table I as this metric was not explicitly reported by authors in several cases. Where it can be compared (see [8], [10], [12]), the previous approaches have outperformed the false positive rate of 2.85 per minute. A possible reason for the lack of reporting on this performance metric in previous approaches is the decision to use frame-level analysis of cough events as opposed to event-level analysis.

An important point that should be noted about our proposed approach is that the algorithm does not require a large amount of training data or a training stage for a neural network, for example. All of the approaches listed in Table I, apart from one, require a training step for the algorithm to detect coughs on a frame-level or event-level. The proposed algorithm here leverages on the repeating nature of the cough sound in the ambulatory recording and does not require an intensive training stage with a large dataset.

V. Conclusions

Building on an approach used for automatically transcribing drum sounds, the proposed algorithm uses ISA for detecting the presence of an individual's cough within audio recordings, reducing the human effort required by medical staff to identify cough events within long recordings.

An average true positive ratio of $76\,\%$, and a false positive rate of 2.85 false positives per minute was achieved, with a significant reduction in the duration of signals. This algorithm's ability to detect cough sounds is evident, and a reduction in false positives will improve the overall performance. A solution to consider in future work utilises multiple analysis windows during the SVD stage, taking advantage of the varying spectral characteristics observed during different phases of a single cough event. Also planned is a more detailed comparison against other state-of-the-art cough-detection approaches.

Acknowledgment

This research is funded by Technological University Dublin's Fiosraigh Scholarship Programme, and in part by Science Foundation Ireland (SFI) (12/RC/2289_P2), and the European Regional Development Fund (13/RC/2077).

References

[1] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[2] S. M. Schappert, "National ambulatory medical care survey: 1991 summary." *Vital and health statistics. Series 13, Data from the National Health Survey*, no. 116, pp. 1–110, 1994.

[3] R. S. Irwin, L.-P. Boulet, M. M. Cloutier, R. Fuller, P. M. Gold, V. Hoffstein, A. J. Ing, F. D. McCool, P. O'Byrne, R. H. Poe *et al.*, "Managing cough as a defense mechanism and as a symptom: a consensus panel report of the american college of chest physicians," *Chest*, vol. 114, no. 2, pp. 133S–181S, 1998.

[4] J. Smith, "Ambulatory methods for recording cough," *Pulmonary pharmacology & therapeutics*, vol. 20, no. 4, pp. 313–318, 2007.

[5] C. T. French, R. S. Irwin, K. E. Fletcher, and T. M. Adams, "Evaluation of a cough-specific quality-of-life questionnaire," *Chest*, vol. 121, no. 4, pp. 1123–1131, 2002.

[6] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19)," *International Journal of Surgery*, 2020.

[7] S. J. Barry, A. D. Dane, A. H. Morice, and A. D. Walmsley, "The automatic recognition and counting of cough," *Cough*, vol. 2, no. 1, p. 8, 2006.

[8] S. Matos, S. S. Birring, I. D. Pavord, and H. Evans, "Detection of cough signals in continuous audio recordings using hidden markov models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1078–1083, 2006.

[9] S. Birring, B. Prudon, A. Carr, S. Singh, M. Morgan, and I. Pavord, "Development of a symptom specific health status measure for patients with chronic cough: Leicester cough questionnaire (lcq)," *Thorax*, vol. 58, no. 4, pp. 339–343, 2003.

[10] S. Birring, T. Fleming, S. Matos, A. Raj, D. Evans, and I. Pavord, "The leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough," *European Respiratory Journal*, vol. 31, no. 5, pp. 1013–1018, 2008.

[11] A. Barton, P. Gaydecki, K. Holt, and J. A. Smith, "Data reduction for cough studies using distribution of audio frequency content," *Cough*, vol. 8, no. 1, p. 12, 2012.

[12] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. N. Patel, "Accurate and privacy preserving cough sensing using a low-cost microphone," in *Proceedings of the 13th international conference on Ubiquitous computing*, 2011, pp. 375–384.

[13] 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.

[14] 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.

[15] J. Amoh and K. Odame, "Deep neural networks for identifying cough sounds," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 5, pp. 1003–1011, 2016.

[16] M. You, Z. Liu, C. Chen, J. Liu, X.-H. Xu, and Z.-M. Qiu, "Cough detection by ensembling multiple frequency subband features," *Biomedical Signal Processing and Control*, vol. 33, pp. 132–140, 2017.

[17] W. Thorpe, M. Kurver, G. King, and C. Salome, "Acoustic analysis of cough," in *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*. IEEE, 2001, pp. 391–394.

[18] D. FitzGerald, "Sub-band independent subspace analysis for drum transcription," in *International Conference on Digital Audio Effects, Germany, 2002*, 2002.

[19] M. A. Casey and A. Westner, "Separation of mixed audio sources by independent subspace analysis." in *ICMC*, 2000, pp. 154–161.

[20] P. Leamy, "paulleamy/isa-cough-detection," Apr 2020. [Online]. Available: https://github.com/paulleamy/ISA-cough-detection.git

[21] C. Uhle, C. Dittmar, and T. Sporer, "Extraction of drum tracks from polyphonic music using independent subspace analysis," in *Proc. ICA*, 2003, pp. 843–847.

[22] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[23] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[24] L. Kvapilova, V. Boza, P. Dubec, M. Majernik, J. Bogar, J. Jamison, J. C. Goldsack, D. J. Kimmel, and D. R. Karlin, "Continuous sound collection using smartphones and machine learning to measure cough," *Digital Biomarkers*, vol. 3, no. 3, pp. 166–175, 2019.

[25] E. Benetos, G. Lafay, and M. Lagrange, "DCASE2016 task 2 baseline," DCASE2016 Challenge, Tech. Rep., September 2016.

[26] C. Bilen, G. Ferroni, F. Tuveri, J. Azcarreta, and S. Krstulovic, "A framework for the robust evaluation of sound event detection," *arXiv preprint arXiv:1910.08440*, 2019.

[27] P. Leamy, D. Berry, T. Burke, and D. Dorran, "Re-annotation of cough events in the AMI corpus," in *2019 30th Irish Signals and Systems Conference (ISSC)*. IEEE, 2019, pp. 1–5.