2019-06-17

Re-Annotation of Cough Events in the AMI Corpus

Paul Leamy
Technological University Dublin, paul.leamy@tudublin.ie

Damon Berry
Technological University Dublin

David Dorran
Technological University Dublin

Ted Burke
Technological University Dublin, ted.burke@tudubln.ie

Recommended Citation

Leamy, P. et al. (2019) Re-annotation of cough events in the AMI corpus, *30th Irish Signals and Systems Conference (ISSC), Maynooth, Ireland, 2019, pp. 1-5, doi: 10.1109/ISSC.2019.8904929*

Re-annotation of cough events in the AMI corpus

Paul Leamy
Biomedical Research Group
Technological University Dublin
Dublin, Ireland
paul.leamy@mydit.ie

Ted Burke
Biomedical Research Group
Technological University Dublin
Dublin, Ireland
ted.burke@dit.ie

Damon Berry
Biomedical Research Group
Technological University Dublin
Dublin, Ireland
damon.berry@dit.ie

David Dorran
Biomedical Research Group
Technological University Dublin
Dublin, Ireland
david.dorran@dit.ie

***Abstract*—Cough sounds act as an important indicator of an individual's physical health, often used by medical professionals in diagnosing a patient's ailments. In recent years progress has been made in the area of automatically detecting cough events and, in certain cases, automatically identifying the ailment associated with a particular cough sound. Ethical and sensitivity issues associated with audio recordings of coughs makes it more difficult for this data to be made publicly available. However, without the public availability of a reliable database of cough sounds, developments in the area of audio event detection are likely to be hampered. The purpose of this paper is to spread awareness of a database containing a large amount of naturally occurring cough sounds that can be used for the implementation, evaluation, and comparison of new machine learning algorithms that allow for audio event detection associated with cough sounds. Using a purpose built GUI designed in MATLAB, the re-annotation procedure followed a reusable methodology that allowed for quick and efficient importing and marking of audio signals, resulting in a re-annotated version of the Augmented Multi-party Interaction (AMI) corpus' cough location annotations, with 1369 individual cough events. All cough annotations and the re-annotation tool are made available for download and public use.**

***Index Terms*—cough event detection, acoustic event detection, AMI Corpus, manual annotation, cough database**

I. Introduction

The information contained within the sounds produced by the human vocal tract such as speech and other vocalisations (such as moaning, sighing, coughing, etc.) present an opportunity to facilitate the remote and non-contact monitoring of an individual's mental state and other health-related issues [1]. Coughing is a common symptom for which patients seek medical advice [2], [3], but the automatic analysis of cough recordings for their identification is difficult due to the random nature between different cough sounds [4].

Recent approaches to analysing coughs have attempted to diagnose pulmonary diseases using feature extraction [5] and to characterise the relationship between cough strength and cough sound [6]. In [7] a preliminary analysis of cough sounds was carried out to extract features that may be useful for categorising cough as *ailment* or *healthy*, and in the process a collection of coughs were collected to create a "Cough Speech Sound" database, containing 110 cough sounds collected from 30 speakers. This data was also marked up with information related to the speaker's name, age, sex, mother tongue, place, and health condition (if any). A substantial amount of cough recordings were gathered in [8], 1967 cough events in total, but all of these recordings come from chest wall recordings. The advantages offered by non-invasive monitoring are removed if equipment must be attached to a person to facilitate monitoring. While all of the databases mentioned so far display merit in terms of the number of cough events, healthy cough, and ailment coughs, one problem still remains. The data is not currently made available for public use. This may be due to ethical considerations when handling sensitive data from patients or because consent was not received from the people whose coughs are present in the recordings.

In some cases, data collected by researchers for cough analysis has been made available to the public. In [9], 9 cough sounds from common respiratory diseases were recorded, where healthcare professionals were required to make a diagnosis about the ailment each patient had based on the sound of the cough. These recordings are available to download in mp3 format, but in the case of designing an automatic cough event detection algorithm, the size of this dataset is not sufficient for training purposes. Alternatively, data sources can come from openly available data. Whooping cough sounds were used from publicly available sources in [10], where links to 38 audio recordings are available in the paper, but these recordings are not properly marked or annotated. The duration of each recording is included, although the number of coughs and start/end times in the corpus is not specified. Fortunately, the Augmented Multi-party Interaction (AMI) corpus [11], [12] overcomes the issues highlighted so far. The AMI corpus data is publicly available under the Creative Commons Attribution 4.0 license agreement, and comes with annotations that specify the position of 1116 cough events present in the audio recordings. Although in Section II-B of this paper, issues concerning the supplied annotations are highlighted.

The layout of the rest of this paper is as follows. In Section II-A an overview of the AMI corpus presents the key points in relation to the data and annotations available with this publicly available corpus. Section II-B discusses the issues with the original annotations regarding the time-stamped cough sound events in the audio recordings in the corpus. The methodology to address these issues with current annotations is outlined in Section II-C, followed by the results and conclusion in Section III and Section IV respectively.

II. Methodology

A. The AMI corpus

The AMI corpus contains 100 hours of time-aligned recordings from a range of synchronised signals including audio and video signals amongst others. Meetings were recorded in which multiple participants act out scenarios representative of realistic meetings. Of the many available annotations included in this corpus, one subset of the annotations is words and non-vocal sounds e.g. laughter, coughing, etc. which are time-stamped for individual speakers in each meeting and role play scenario. In each meeting recording 3-5 participants are present. In terms of the audio signals, recordings were captured from multiple microphone locations including,

- Headset microphone (one per participant)
- Lapel microphone (one per participant)
- Omni-directional microphone arrays (near and far mounted)

with all the individual recordings synchronised.

The AMI corpus is supplied with a set of manual annotations, where a wide range of events for individual participants at each meeting are labelled and time-stamped. The word annotations contain information about individual speakers in an XML file. Information corresponding to the meeting code, speaker I.D., word number, start time, end time, and spoken word are located in these files. An example of an annotation for a single word is shown below for the word “to”,

```
<w nite:id="EN2001a.A.words783" starttime="2860.29"
    endtime="2860.74">to</w>
```

where this word started at time 2860.29 s and ended at time 2860.74 s during meeting EN2001a, and was spoken by speaker A. Vocal sounds that do not correspond to spoken words are also marked up in these annotations. These vocal sounds include what the annotators consider to be cough sounds. An example of an annotation that marks a vocal sound as a “cough” is included below,

```
<vocalsound nite:id="EN2001a.A.words789" starttime="
    3099.54" endtime="3099.97" type="cough"/>
```

where a cough vocal sound started at time 3099.54 s and ended at time 3099.97 s during meeting EN2001a, and was spoken by speaker A.

Following an initial analysis of the annotations provided, it was found that there was a total 1116 events labelled as “cough” vocal sounds in the AMI corpus annotations. A more in-depth analysis of these annotations is provided in the Section II-B, which highlights multiple issues concerning the manual annotations provided with AMI corpus.

B. Issues with original annotations

The annotations that are currently provided with the AMI corpus include the start and end times of cough events for individual speakers in each meeting. In Fig. 1, a histogram is used to display the distribution of the duration of all of the annotated cough times from the original AMI corpus annotations. In Fig. 1, two main issues can observed. First, a large peak at 0 s can be seen. The duration values are calculated as the difference between the end time and start time for each cough event. The 0 s duration values suggest that the start and end times entered into the original XML file are the same for a large number of cough events. Also, a substantial number of the cough event annotations resulted in a duration range between 1 s and 3 s, which is considerably longer than a typical single cough event. The two issues are highlighted in Fig. 2 and Fig. 3.

In Table I, a summary of the cough duration values is provided. Following an analysis of the original annotations, it was found that in total, 333 of the 1116 cough event annotations were marked with a duration of 0 s, and 2 instances with a negative duration. The remaining number of cough event annotations were marked with a start time and end time resulting in a positive duration.

TABLE I
Distribution of cough events

	Cough durations		
	$> 0s$	$= 0s$	$< 0s$
# Cough events	781	333	2

Instances where a duration of 0 s for a cough event were noted and analysed to determine which phase of the cough these annotations were referring to i.e. the explosive, intermediate, or ending phase. The result of this analysis was that 0 s

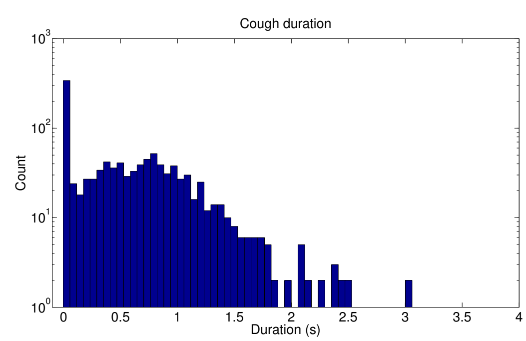

Fig. 1. Distribution of cough duration based on the original AMI corpus annotations. Following an initial analysis of the annotations, a large number of the time-stamped cough events had been annotated with start and end times that were identical resulting in a duration of 0 s. Also a number of cough events had been marked with start and end times resulting in an unrealistically long duration time in the range of 1 s to 3 s. Note the logarithmic scale on the count axis.

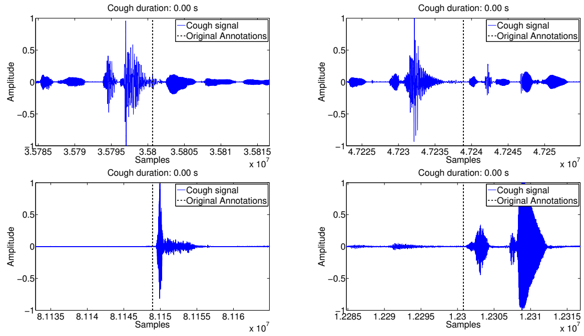

Fig. 2. Examples of instances where start and end time annotations resulted in 0 s duration for a cough event. In these four examples taken from the AMI corpus annotations, inconsistencies in the marking of either the start or end times of the cough events is present. Original annotations are displayed as vertical dashed lines. These inconsistencies meant that using the provided annotations could not lead to reliable marking of cough events. For clarity, cough recordings shown in these figures are taken from the individuals' headset microphone recordings, as a result background noise is not as apparent as in the microphone array recordings.

cough events were returning inconsistent positions in relation to the start phase, middle phase, or end phase of the coughs. In Fig. 2 four examples of 0 s duration cough events are shown. In these examples, it can be seen that different phases of the cough event have been marked with the original annotations. In the top two figures, the end of a cough event has been marked, while in the bottom two figures the beginning phases of the cough events have been marked. Following further investigation into the original annotations for the cough events, it was found that the problem of inconsistent marking of start and end times was not specific to instances where 0 s cough events had been labelled. This problem occurred in the majority of the cough annotations supplied with the AMI corpus.

In Fig. 1 it can also be seen that there are a number of coughs with a duration in excess of 1 s are present. These cough events were also analysed and examples of these are included in Fig. 3. All four examples of long duration cough events included in Fig. 3 contain more than one single cough event. The issues highlighted in Fig. 2 regarding inconsistent and inaccurate marking of start and end times are also present in the cough events shown in Fig. 3.

While the AMI corpus has significant potential to aid in the design and test of automatic cough detection algorithms, the inaccuracies and inconsistencies present in the existing corpus annotations limit this potential in its current state. To deal with these issues the authors have re-annotated the coughs using the procedure described in the following section.

C. Re-annotation procedure

The annotations supplied with the AMI corpus, that time-stamp cough events from the individual speakers could potentially aid the development, implementation, and testing of algorithms concerned with the detection of cough audio events. Unfortunately, the issues and inconsistencies highlighted in the previous section currently render this information unreliable as an accurate indicator of *single* cough events in the audio recordings from the AMI corpus. To resolve this issue, the authors manually re-annotated each cough event listed with the following features:

- Start/onset time (in samples)
- End time (in samples)
- Where multiple coughs occur in rapid succession, they are annotated as individual cough events, unlike in the original annotations

where a single cough event relates to a single expulsion of air from a person's respiratory system (the existing annotations often indicate a quick succession of coughs as a single cough event). The task of importing and analysing the audio recordings, and recording the new time-stamps was carried out using MATLAB. For this task, a GUI was created that allows the annotator to perform the tasks below.

- Import original time stamps
- Import the relevant sections of audio
- Play back the two audio signals independently
- View the time-domain cough signal
- Assess the accuracy of the original time-stamps
- Create new time stamps for each cough event where necessary
- Save to file

An outline of the GUI is shown in Fig. 4. Room recordings and the relevant playback control buttons are situated under the axes on the left, while the headset recording and playback controls are under the axes on the right. Time stamps can be recorded directly from either axes to allow new time stamps to be stored. The individual playback controls allow the annotator to properly identify the start and end times of each single cough event in the audio recordings.

In relation to determining which points of the audio recording are marked for the start and end times of a cough event, the reader is referred to the example annotation included in Fig. 5.

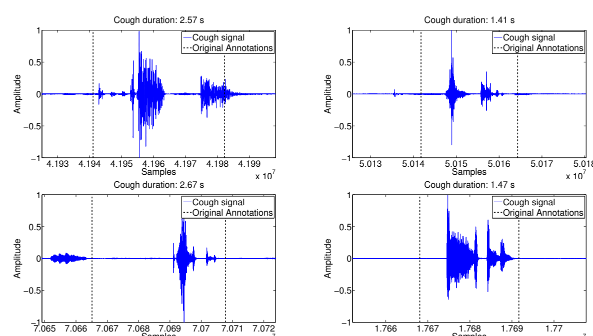

Fig. 3. Examples of instances where the duration for a cough event is longer than 1 second. In these four examples, an issues can be seen regarding the provided time-stamps in the original annotations. Cough events marked with start and end times resulting in an unusually long duration have been observed to contain cough events with multiple single coughs. Original annotations are displayed as vertical dashed lines. For clarity, cough recordings shown in these figures are taken from the individuals' headset microphone recordings, as a result background noise is not as apparent as in the microphone array recordings.

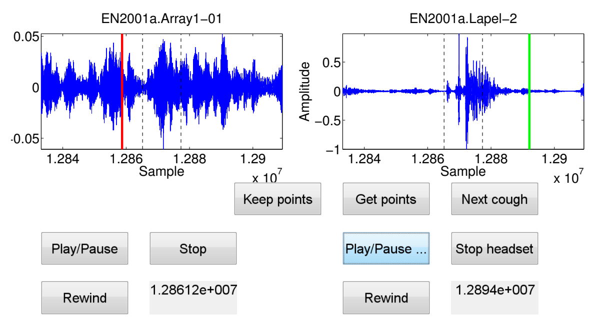

Fig. 4. The GUI created to carry out the process of annotating the cough events in the AMI corpus. This tool allows the user to import audio signals from room and headset recordings separately. Audio signals can be played back, paused, and rewound to accurately identify the point where a cough event starts and ends. Playback markers for each signal can be seen in red (left: room) and green (right: headset). The original time-stamps are included and show up as dashed vertical black lines. Multiple new time-stamps can be created if the recording contains more than one cough event.

We define the start time as the beginning of the explosive phase of a cough. We define the end time as the point which the intermediate phase or voiced phase of a cough decays to silence. In the case where successive coughs occur and the sound does not decay to silence, the transition between the decaying cough and explosive section of the next cough is used as the end time of the decaying cough and the start time of the new cough sound.

III. Results

The purpose of the new annotation procedure described in Section II was to consistently mark the start and end times of single cough events in the AMI corpus. Fig. 1 highlights two issues with the original annotations provided with the AMI corpus. The first issue was concerned with the annotations that suggested a high proportion of the total number of originally annotated cough events had a duration of 0 s, a total of 333 instances. The second issue was concerned with cases of excessively long cough durations, where in Fig. 3 it is observed that multiple cough events were being marked with a single start and end time.

Following this annotation procedure, a new set of time stamps corresponding to the start and end times of single cough events was produced, and new time stamps for the cases where multiple single coughs were marked as a single cough event were produced. The total number of single cough events before and after the new annotation procedure are included in Table II. The number of marked single cough events has increased from 1116 to 1369, a 22.67% increase in the number of cough events relative to the original annotations.

TABLE II
Distribution of Cough Events

Number of cough events	
Original annotations	*New annotations*
1116	1369

In Fig. 6 the distribution of cough times for all the newly annotated events is shown. This can be compared directly with the previous distribution of cough times as shown in Fig. 1, noting the logarithmic scale used in Fig. 1 and the linear scale used in Fig. 6.

In Fig. 6, the distribution of the cough durations based on the new start and end times is shown. It can be seen that the two issues have been resolved. In the case of the first issue, the number of 0 s cough events has reduced to 0, where all annotations accurately mark the start and end of the cough events, as described in Fig. 5. Also, the range of the duration values now resides in more realistic boundaries with only a small number of events residing in the regions above 1 s. The cough sounds with the longer duration were analysed separately, and it was noted that these events have a "wheezy" characteristic associated with them. In these cases, the initial explosive section of the cough sound is followed by an extended sustained expulsion of air, resulting in the longer duration.

The new annotations are available as a set of MAT-files

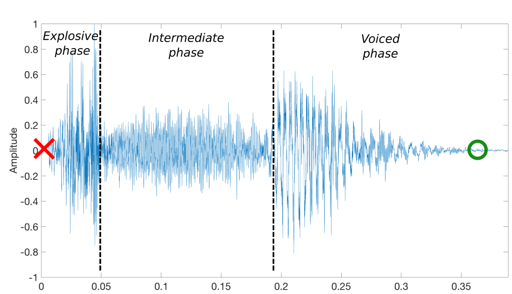

Fig. 5. Description of the criteria for determining the start and end times to be used while annotating the cough events in the audio recordings from the AMI corpus. The red **"x"** marker is located before the beginning of the onset of the cough (before the explosive section) and this marks the beginning of the cough. The green **"o"** marker located at the end of the cough event marks the end time of the cough sound.

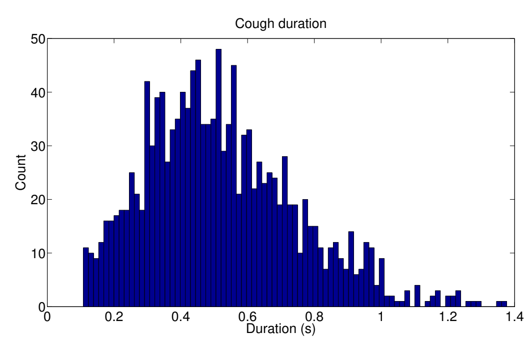

Fig. 6. Distribution of cough duration based on the new annotations. Following the steps that were described in Section II the amount of 0 s duration cough events has reduced dramatically. Also, multiple cough events have been identified and annotated to represent the single cough events resulting in a reduction in the number of unusually long cough events. Note the linear scale on the count axis.

```
CoughAnnotations
 ├── EN2002a
 │    ├── ``startTimes''
 │    ├── ``endTimes''
 │    └── ``speaker''
 ├── EN2002b
 │    ├── ``startTimes''
 │    ├── ``endTimes''
 │    └── ``speaker''
 └── ⋮
```

Fig. 7. Structure of the file containing the new annotations. Each of the 152 meeting files contains the start and end times for each cough event and speaker contains an ID letter that identifies the individual who produced the cough event.

with one MAT-file available for each meeting recording where cough events are present. Each individual MAT-file contains the start times, end times, and speaker identifier letter for a set of cough events. The inclusion of the speaker identifier letter allows the individual recording to be identified. The annotations and GUI are available for download as ZIP files from GitHub using the URL in [13], while the file structure of the annotations is outlined in Fig. 7.

IV. Conclusion

The procedure of re-annotating the cough events contained in AMI corpus is presented in this paper. The justification for the re-annotation came following an initial analysis of the provided annotations, where it was found that a number of issues existed relating to the original start and end times of the cough events in the audio recordings. Inconsistencies with the original annotations of the cough events came in the form of both incorrect start and end times, and successive cough events being annotated as a single cough event. These inconsistencies lowered the potential for the AMI corpus to become a reliable and reusable source of cough recordings. In the area of cough sound analysis and detection, where publicly available data with annotations is difficult to obtain due to ethical considerations, the presence of a reliably annotated database will be a useful addition, progressing the development and testing of new audio event detection algorithms concerned with cough sounds, particularly in the area of machine learning.

Following the re-annotation procedure, a total of 1369 individual cough events, up from 1116, are now correctly annotated. It should be noted that the re-annotation procedure presented in this paper used the manual annotations provided by the AMI corpus to identify and accurately re-annotate the cough events. Possible future work concerned with analysing the full extent of the 100 hours of audio recordings contained in the AMI corpus may result in the accurate detection of new cough sounds, ones not originally recorded by the manual annotators who produced the 1116 cough events. As a result, the actual number of cough events labelled in these recordings may rise again, following further analysis.

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] S. M. Schappert and C. W. Burt, "Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United states, 2001-02." *Vital and Health Statistics. Series 13, Data from the National Health Survey*, no. 159, pp. 1–66, 2006.

[4] J. Schrder, J. Anemiiller, and S. Goetze, "Classification of human cough signals using spectro-temporal gabor filterbank features," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6455–6459.

[5] C. Infante, D. Chamberlain, R. Fletcher, Y. Thorat, and R. Kodgule, "Use of cough sounds for diagnosis and screening of pulmonary disease," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2017, pp. 1–10.

[6] K. K. Lee, S. Matos, K. Ward, G. F. Rafferty, J. Moxham, D. H. Evans, and S. S. Birring, "Sound: a non-invasive measure of cough intensity," *BMJ open respiratory research*, vol. 4, no. 1, p. e000178, 2017.

[7] V. P. Singh, J. M. S. Rohith, and V. K. Mittal, "Preliminary analysis of cough sounds," in *2015 Annual IEEE India Conference (INDICON)*, Dec 2015, pp. 1–6.

[8] B. Rocha, L. Mendes, R. Couceiro, J. Henriques, P. Carvalho, and R. P. Paiva, "Detection of explosive cough events in audio recordings by internal sound analysis," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 2761–2764.

[9] J. A. Smith, H. L. Ashurst, S. Jack, A. A. Woodcock, and J. E. Earis, "The description of cough sounds by healthcare professionals," *Cough*, vol. 2, no. 1, p. 1, 2006.

[10] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PloS one*, vol. 11, no. 9, p. e0162128, 2016.

[11] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus," *Language Resources and Evaluation*, vol. 41, no. 2, pp. 181–190, 2007.

[12] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal *et al.*, "The ami meeting corpus: A pre-announcement," in *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 2005, pp. 28–39.

[13] P. Leamy, "paulleamy/ami-cough-annotations," Mar 2019. [Online]. Available: https://github.com/paulleamy/AMI-Cough-Annotations.git