様　式　C－19、F－19－1、Z－19　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　６ 月 ２０ 日現在

機関番号：２１６０２
研究種目：基盤研究(C)（一般）
研究期間：2020～2023
課題番号：２０Ｋ１２０８０

研究課題名（和文）咳の自動監視・計測・サービスシステムの構築

研究課題名（英文）Development of automatic cough monitoring, measurement and service system

研究代表者
ＭＡＲＫＯＶ　Ｋ（Markov, Konstantin）

会津大学・コンピュータ理工学部・教授

研究者番号：８０３９４９９８

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：この研究では、音声録音における咳イベントの効率的な検出とモニタリングのための新しいディープラーニング システムを提案します。私たちのアプローチは、いくつかの注目すべき点でユニークです。1) 検出モデルのトレーニングに教師と生徒のフレームワークを組み込んでいます。2) MFCC やメル スペクトログラムなどの特定のオーディオ機能に頼るのではなく、エンドツーエンドのシステムは生のオーディオ信号を直接入力として受け取り、咳の境界タイミングを出力します。LCM などのシステムと比較した咳モニタリング評価では、最も低いエラーを達成することで、私たちの方法の優位性が実証されています。

研究成果の学術的意義や社会的意義
The proposed cough detection and monitoring approach has the potential to impact various domains beyond respiratory health, such as urban planning, environmental monitoring, and animal welfare. The methodology is independent and adaptable to any sound event detection task requiring counting events.

研究成果の概要（英文）：In this research, we propose a novel deep learning system for the efficient detection and monitoring of cough events in audio recordings. First, we perform voice activity detection to eliminate audio silences. Next, we employ a cough classification to identify the presence of cough. Finally, we implement cough event detection using a high-performance classification-regression fusion method. Our approach is unique in several notable ways: 1) we incorporate a teacher-student framework for the training of our detection model, 2) instead of relying on specific audio features such as MFCC or Mel Spectrogram, our end-to-end system takes the raw audio signal directly as input and outputs the cough boundary timings, 3) the proposed method is general enough to be used for various other sound event monitoring tasks. The comparative cough monitoring evaluation of our approach against systems such as the Leicester Cough Monitor demonstrates our method's superiority by achieving the lowest error.

研究分野：audio event detection

キーワード：cough monitoring　health informatics　audio event detection

１．研究開始当初の背景

Cough serves as a crucial diagnostic indicator, being a common symptom related to many respiratory diseases, including chronic obstructive pulmonary disease, asthma, and, not to mention, COVID-19 and its variants. In practice, cough frequency monitoring enables doctors to track the disease severity and verify the effectiveness of the treatment. However, traditional cough diagnosis typically relies on subjective reports of patients, such as the Leicester Cough Questionnaire or the Cough-Specific Quality of Life Questionnaire which may not provide accurate information. In contrast, an automatic and precise cough monitoring system will not only better assist professionals with disease diagnosis and treatment but also benefit the development of new healthcare services.

The fundamental component of a cough monitoring system is cough detection. It falls under the category of sound event detection tasks, intending to identify and pinpoint the target event and its corresponding temporal location within the audio signal. The typical approach for sound event detection begins with the extraction of audio features, followed by the utilization of a classifier to determine whether the target event of interest is present within the input signal. When the goal is to simply assess the presence of a specific sound event within a short audio segment, the result of this step is sufficient and the task is commonly known as segment-based sound event detection. In other scenarios where identifying the boundaries of the target events is essential, an additional step is required, where the event's onset and offset timing have to be determined. This is usually achieved by performing frame-wise audio tagging. However, for comprehensive cough monitoring, this approach suffers from low performance issues, primarily due to missed boundaries in multiple continuous cough events. Cough detection inherits the challenges of sound event detection along with its own. Unlike speech recognition, sound event detection is, in practice, task-specific, as there is no established ontology to universally define sound events. In addition, because of the additive nature of sounds, when multiple sound sources are active at the same time, the detection task becomes more challenging. Furthermore, the distance between the sound source and the recording device, as well as the biases introduced by different types of recording devices, are also adding difficulty to the task. The challenges for detecting cough itself include the large data variation and the difficulty to identify a single cough boundaries in continuous coughs, which is crucial for cough monitoring. The characteristics of cough sounds depend on many factors, such as differences in the airway, lungs, and vocal cords. Regardless of the cause of these differences, they lead to different kinds of cough sounds. For example, coughs from obstructed airways, like bronchiectasis or asthma patients, tend to have stronger energy and shorter duration, while coughs from interstitial lung disease patients have the longest duration, and coughs from healthy airways and lungs tend to have a higher frequency. When the final objective is to monitor the frequency of coughs, the system has to be able to identify a single cough boundaries, especially for cases involving multiple continuous coughs. For instance, when two continuous coughs are present, the traditional frame tagging approach may detect one cough, while there are two coughs. This may occur because the interval between continuous coughs is too short for such an approach to work reliably.

２．研究の目的

To address the aforementioned challenges and to provide a comprehensive and automated solution, we propose a novel end-to-end deep neural network (DNN) based system which we call the 'Neural Cough Counter'. The key contributions of this study are as follows:

・Combination of transfer learning with distillation training for improved frame-level cough detection: We harness the capabilities of transfer learning by fine-tuning a pre-trained speech model for feature extraction. This process in combination with distillation training enhances the quality and robustness of the features used for cough detection, making our model more powerful and reliable.

・Innovative Fusion Model for Cough Detection: Our design of a regression and classification fusion model offers an innovative approach to cough boundary detection. This fusion model successfully overcomes the challenges inherent in traditional frame-tagging approaches. It not only enhances the identification of single cough events within continuous cough sequences but also provides a foundation for precise cough monitoring.

・Application to a Range of Sound Event Detection Tasks: While our primary focus is on cough detection and monitoring, our proposed framework can be potentially applied to a wide range of sound event monitoring tasks. By leveraging deep learning techniques, our model showcases adaptability that could extend to various sound event detection applications.

By incorporating these innovations, our study paves the way for more accurate and automated healthcare diagnostics and monitoring systems, contributing to health-related research.

３．研究の方法

An overview of our cough detection and monitoring system is presented in Fig. 1. It takes waveform audio as input and outputs the time boundaries of the detected coughs according to the following steps:

(1) Given an audio recording, voice activity detection is first performed to remove the silent parts of the audio. This way, the amount of data for further processing is greatly reduced. This step results in a set of audio segments containing speech, cough, noise, etc.

(2) Next, we perform a classification of the segments from the previous step. Here we aim to identify the segments that contain cough events. This step improves the performance of the system, as it removes most of the non-cough segments and reduces the potential false positives in the cough detection step.

(3) Finally, we run our cough detection model on each cough segment identified by the previous step and locate the beginning and ending time of every cough event.

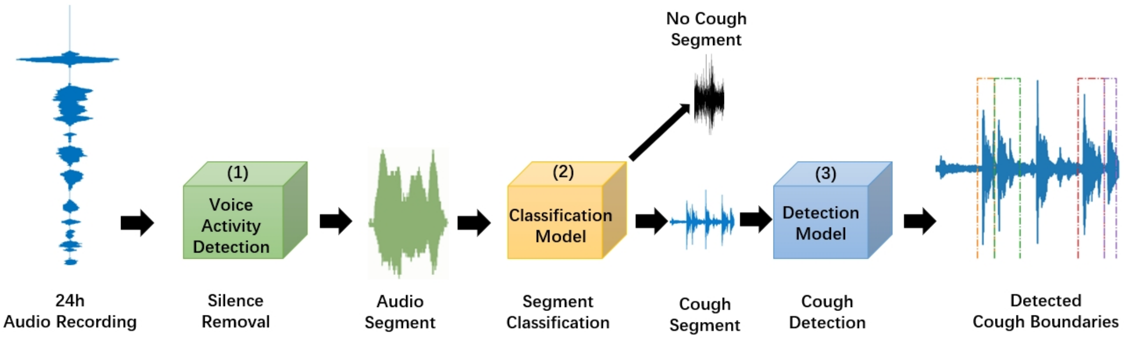

Fig.1 The pipeline of the proposed system in this study follows the steps below: (1) Voice Activity Detection is firstly performed on 24 hours of audio recording for silence removal, resulting in a set of audio segments. (2) Segment Classification is subsequently performed to obtain the cough segments. (3) Cough detection is finally carried out on the cough segments to obtain the cough boundaries.

４．研究成果

Through extensive experiments and analysis, we demonstrated the superiority of our classification model compared to the baseline CRNN model in terms of various evaluation metrics. Our model exhibited enhanced performance in accurately identifying and distinguishing cough segments, showcasing the effectiveness of using pre-trained models and fine-tuning them on target cough data. Furthermore, we explored the use of knowledge distillation to train a student model with improved frame level classification performance resulting in more accurate cough event boundaries detection. In addition, to overcome the limitations of existing studies in cough detection and monitoring, we proposed an innovative regression and classification fusion approach. This approach enabled us to train a cough detection model, which exhibits remarkable capabilities in detecting the boundaries of individual cough events within continuous cough instances. We rigorously evaluated our model on both the in-house and public datasets, demonstrating its robustness and generalizability. The strong performance of our model on the public dataset highlights its potential for real-world applications and its ability to handle diverse cough data from different sources. Finally, we conducted experiments to compare the performance of different models in cough monitoring. The results highlighted the exceptional performance of our model, surpassing the system baseline and other models such as the LCM, XGBoost , and the Universal approach in estimating the hourly cough counts with much lower hourly average sMAPE values shown in the following table:

			Hourly symmetric Mean Absolute Percentage Error ($sMAPE^{100}$) on the In-House dataset					
Patient	Cough Counts	Hours	Detection Model	Frame Model	CRNN	LCM [30]	XGBoost [31]	Universal(tuned) [32]
1	772	24	**2.02**	4.94	3.26	37.92	46.05	32.58
3	1298	24	**3.56**	6.97	5.05	25.62	42.29	28.51
4	475	6	**8.96**	9.29	19.51	36.56	48.24	36.94
5	448	24	**11.49**	12.86	15.25	12.11	40.51	43.44
6	623	24	**7.94**	13.45	16.03	49.77	52.78	50.00
7	234	6	**7.57**	10.67	21.96	40.33	19.33	34.91
8	749	1	**11.21**	24.11	33.16	59.08	37.56	39.61
9	266	18	**7.55**	11.09	12.69	25.80	41.87	45.70
10	229	24	**16.07**	17.89	18.28	61.26	43.07	49.81
Average Hourly $sMAPE^{100}$			**8.48**	12.36	16.13	38.72	41.30	40.17

5．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　1件）

1．著者名	4．巻
Z.Feng, K. Markov, J.Saito, T. Matsui	0
2．論文標題	**5．発行年**
Neural Cough Counter: A Novel Deep Learning Approach for Cough Detection and Monitoring	2024年
3．雑誌名	**6．最初と最後の頁**
IEEE Access (Accepted for publication)	0-0
掲載論文のDOI（デジタルオブジェクト識別子）	**査読の有無**
なし	有
オープンアクセス	**国際共著**
オープンアクセスとしている（また、その予定である）	該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	松井　知子 (Matsui Tomoko) (10370090)	統計数理研究所・モデリング研究系・教授 (62603)	
研究分担者	齋藤　純平 (Saito Jimpei) (50332929)	福島県立医科大学・医学部・講師 (21601)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関