

スマートウォッチを用いた咳嗽好発時間帯記録システムの構築

山名 風太^{1,a)} 奥野 拓^{1,b)}

概要: 咳嗽の発生時刻を記録し、咳嗽が多く発生する時間帯（咳嗽好発時間帯）を知ることは原因疾患の特定に有効である。しかし現状の医療機関で用いられている方法は患者への負担が大きい。そこで本研究では、患者の負担を軽減するために、装着時の違和感が小さいスマートウォッチを用いて咳嗽の発生時刻を自動的に記録するシステムを構築する。システムはスマートウォッチを用いて患者の音声を録音し、咳嗽検出アルゴリズムによって咳嗽と非咳嗽を分類する。本研究ではより精度が高いアルゴリズムを選定するため、実際の患者の咳嗽音を用いて二つのアルゴリズムを比較する。評価実験の結果、ランダムフォレストを用いて学習させたモデルを使用したアルゴリズムに比べて、Apple 社が公開している音声分析用のフレームワークである Sound Analysis の方が精度が高いことが分かった。そのため、本研究では咳嗽検出に Sound Analysis を用いる。また実験用のシステムを構築して最大稼働時間を評価した結果、患者の音声を常時録音するには不十分であることが分かった。そのため、録音時の音質を低下させることによって、最大稼働時間を延長する必要がある。

1. はじめに

咳嗽は医療機関への受診理由として最も多い症状の一つである。しかし、咳嗽の原因疾患は非常に多く、特定するのは困難であると言われている。原因疾患を特定する際に有効な指標の一つとして、咳嗽が多く発生する時間帯（以下、咳嗽好発時間帯とする）が挙げられる。その例を図1に示す。喘息とうっ血性心不全は夜中から朝方にかけて、心因性咳嗽と胃食道逆流症は日中に多く咳嗽が発生していることがわかる。このように、原因疾患によって咳嗽好発時間帯が異なるため、診断の際に有効な指標となる。

咳嗽好発時間帯を知るためには、咳嗽の発生を記録する

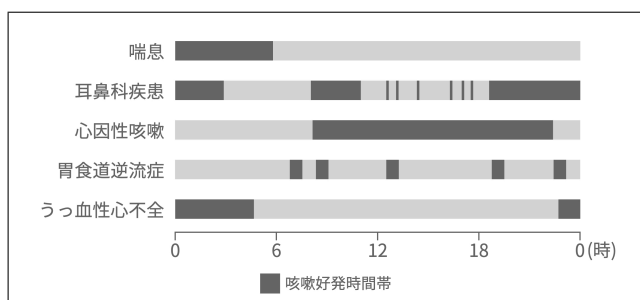


図1 咳嗽好発時間帯の例

必要がある。しかし、咳嗽の発生を記録することは患者への負担が大きいという問題がある。現状の医療機関において用いられている方法は主に二つある。一つ目は、患者自身による記録であるが、この方法は記録に多くの時間と労力が必要になる。二つ目は、専用の機材による記録であるが、この方法は機材装着時の違和感が大きく、日常生活に支障をきたす場合がある。このように、現状の医療機関において用いられている方法は様々な面で患者への負担が大きい。それを軽減することが求められている。

そこで本研究では、患者への負担が少ない咳嗽発生時刻の記録を目的とする。この目的を達成するために、装着時の違和感が少ないスマートウォッチを用いて、咳嗽発生時刻を自動的に記録するシステムを構築する。

2. 関連研究

スマートウォッチを用いた咳嗽検出に関する研究として、Liaqat らの研究が挙げられる [1]。Liaqat らはスマートウォッチを用いて録音した音声の品質を調査するために、それらから咳嗽が検出可能かどうか検証した。具体的には、録音した音声からメル周波数ケプストラム係数（以下、MFCC とする）を抽出し、ランダムフォレストを用いて学習させたモデルを使用し、咳嗽と非咳嗽を分類している。分類の結果は再現率が 91% と非常に高い精度であり、スマートウォッチを用いて録音した音声でも十分に咳嗽を

¹ 公立はこだて未来大学
Future University Hakodate
a) g2119044@fun.ac.jp
b) okuno@fun.ac.jp



図 2 咳嗽発生時刻を自動的に記録するためのシステムの構成

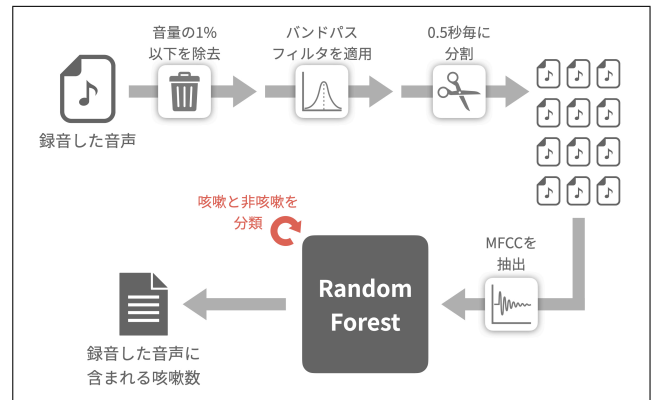


図 4 録音した音声から咳嗽を検出するアルゴリズム

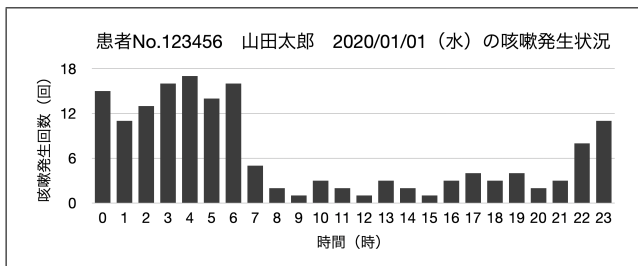


図 3 医師へ提示する咳嗽発生状況のイメージ

検出できることが判明した。しかし、このシステムは録音した音声を全てサーバへ送信しているため、使用時の通信量が膨大になる。また、自分の生活音を第三者に共有することは、患者にとって心理的な負担が大きい。そのため、実際に患者に使用してもらうためには、これらの問題を解決する必要がある。

3. 構築するシステム

本研究で構築するシステムの構成を図 2 に示す。システムはスマートウォッチを用いて患者の音声を録音する。録音は入浴中などスマートウォッチを装着できない場合を除き、常時行う。また検出精度向上のため、録音は非圧縮形式で行う。スマートフォン上で咳嗽を検出するため、録音した音声をスマートフォンへ送信する。これはスマートフォン上で咳嗽を検出した方がスマートウォッチのバッテリー消費量が少なく、より長い時間システムを稼働させられるためである。またメモリの使用量が多すぎると OS によってシステムが強制停止させられる可能性があるため、音声は 30 分ごとに送信する。送信された音声から 4 章で述べるアルゴリズムを用いて咳嗽を検出し、音声の録音開始時刻と終了時刻、音声内に含まれる咳嗽の数をサーバへ送信する。送信されたデータを用いて患者の 1 日における咳嗽発生状況を作成し、図 3 のような形式で医師へ提示する。

4. 咳嗽検出アルゴリズム

本研究では、より精度が高いアルゴリズムを選定するため、二つのアルゴリズムを比較する。一つ目は Liaqat らの

手法を元にしたアルゴリズムである（以下、Liaqat-based とする）。これは Liaqat らが咳嗽の検出に用いている特徴量と学習方法に、独自の前処理を追加したものである。Liaqat らの手法を元にした理由は、再現率が 91% と非常に高かったためである。また独自の前処理を追加した目的は、咳嗽の検出率向上およびバッテリー消費量の削減である。

二つ目は Sound Analysis [2] である。Sound Analysis は、Apple 社が公開している音声分析用のフレームワークである。音声と Create ML [3] を用いて学習させたモデルを入力すると、自動で特徴量の抽出と音声の分類を行う。Create ML は、Apple 社が公開している分類モデル作成用のフレームワークである。学習データに用いる音声、および学習回数の上限と特徴量を抽出する際の重複率を入力すると、自動的に特徴量を抽出し、モデルの学習を行う。これらの特徴によって、システムを実装・保守する際のコストが低くなるため、アルゴリズムの候補として採用した。以下の節では、それぞれのアルゴリズムについて述べる。

4.1 Liaqat-based

Liaqat-based の処理手順を図 4 に示す。以下の項では具体的な処理内容について述べる。

4.1.1 特徴量の抽出

常時録音した音声は時間が長く、そのまま分析すると多くのバッテリーを消費すると考えられる。そのため、まず無音に近い部分を除去し、時間を短くする。本研究では音声ファイル内における最大音量の 1% 以下を無音に近い部分とする。スマートウォッチを用いて録音した音声と、その音声から無音に近い部分を除去した音声の波形を図 5 に示す。振幅を ±1.0 の範囲で正規化しているため、-0.01 ~ 0 および 0 ~ 0.01 に含まれる部分が除去されている。元の音声の時間は 9431 秒であったが、除去によって 71 秒になっており、大幅に短縮できていることがわかる。

無音に近い部分の除去後、バンドパスフィルタ（以下、BPF とする）によって環境音の低減を行う。本研究では咳嗽の特徴がよく出るとされている 100Hz ~ 900Hz [4] 以外を減衰させる。スマートウォッチを用いて録音した音声

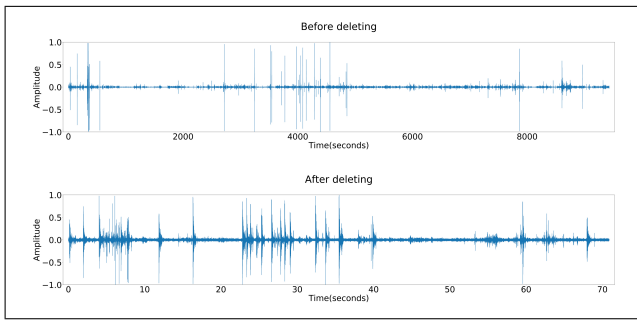


図 5 録音した音声および最大音量の 1%以下を除去した音声の波形

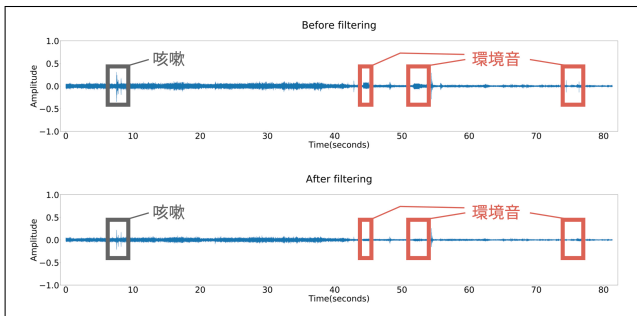


図 6 録音した音声および BPF を適用した音声の波形

と、その音声に BPF を適用した音声の波形を図 6 に示す。図内に示している三つの環境音の振幅が、BPF 適用後は 0 に近い値になっていることがわかる。

BPF の適用後、音声を 0.5 秒ごとに分割し、特徴量として MFCC を抽出する。本研究では対象とする咳嗽のほとんどが 0.5 秒以内であったため、分割後の音声の長さを 0.5 秒とする。また MFCC の次元数は一般的な 12 次元とする。

4.1.2 咳嗽の検出

咳嗽と非咳嗽を分類するモデルを作成するために、ランダムフォレストで学習を行なう。学習データには Apple Watch Series 3 を用いて録音した音声および Freesound [5] から収集した音声から、目視と聴音によって手動で切り出したものを用いる。学習データの内訳は、咳嗽音が 1009 個、日常生活で生じる様々な非咳嗽音が 1009 個の合計 2018 個である。この学習データから 4.1.1 項の方法で特徴量を抽出して学習を行い、学習済みモデルを作成する。咳嗽の検出は、この学習済みモデルを用いて咳嗽と非咳嗽を分類することによって行う。

4.2 Sound Analysis

4.2.1 特徴量の抽出

Sound Analysis は、Create ML を用いて以下の手順で特徴量を抽出する。

- (1) 振幅を ± 1.0 の範囲で正規化する
- (2) 音声のチャンネルが二つ以上ある場合、平均を取り一つに統合する
- (3) 16kHz のサンプリング周波数で再標本化する

- (4) モデルを作成する際に設定する重複率に従い、975ms ごとに音声を分割する
- (5) それぞれの区間にハミング窓を適用する
- (6) FFT を用いてパワースペクトルを算出する
- (7) BPF を用いて 125Hz～7500Hz 以外を減衰させる
- (8) メルフィルタバンクを適用する
- (9) 全ての値に対して自然対数をとる
- (10) VGGish [6] から最後の三つの層を取り除いたものを用いて 12288 次元のベクトルを抽出する

4.2.2 咳嗽の検出

咳嗽と非咳嗽を分類するモデルを作成するために、Create ML を用いて学習を行なう。学習データには 4.1.2 項と同様の音声を用いる。この学習データを Create ML に入力し、学習済みモデルを作成する。なお学習回数の上限と特徴量を抽出する際の重複率に関しては、どちらも標準値である 25 回と 50%とする。咳嗽の検出は、この学習済みモデルを用いて咳嗽と非咳嗽を分類することによって行う。

5. 咳嗽検出アルゴリズムの精度評価

5.1 実験内容

咳嗽検出アルゴリズムの精度および BPF の有用性を検証するために、実験を行なった。BPF の有用性を検証するため、実験には Liaquat-based, Liaquat-based (BPF 未適用), Sound Analysis の三つのアルゴリズムを用いた。Liaquat-based (BPF 未適用) は、4.1.1 項の処理から BPF を適用する処理を除いたアルゴリズムである。テストデータには Apple Watch Series 3 を用いて録音した音声から目視と聴音によって手動で切り出したものを用いた。テストデータの内訳は、咳嗽音が 22 個、非咳嗽音が 46 個の合計 68 個である。このテストデータをそれぞれのアルゴリズムに入力し、精度を評価した。

5.2 結果と考察

録音場所および音声提供者の性別と年齢、ファイルに含まれる咳嗽数と非咳嗽数を表 1 に、それぞれのアルゴリズムの分類結果を表 2, 3, 4 に、それぞれのアルゴリズムの F 値を表 5 に示す。表 2 と表 3 より、Liaquat-based は Liaquat-based (BPF 未適用) に比べて咳嗽の検出率が 4.5pt、非咳嗽の検出率が 31.6pt 低いことが分かった。ま

表 1 録音場所および音声提供者の性別と年齢、ファイルに含まれる咳嗽数と非咳嗽数

場所	性別	年齢	咳嗽数	非咳嗽数
寝室	男性	58	1	0
大学内	男性	58	4	0
研究室	男性	58	2	0
診療室	男性	58	2	0
診療室	女性	22	4	8
診療室	女性	38	9	38

表 2 Liaqat-based の分類結果

		予測値 (個)	
		咳嗽	非咳嗽
実測値 (個)	咳嗽	20	2
	非咳嗽	20	26

表 3 Liaqat-based (BPF 未適用) の分類結果

		予測値 (個)	
		咳嗽	非咳嗽
実測値 (個)	咳嗽	21	1
	非咳嗽	8	38

表 4 Sound Analysis の分類結果

		予測値 (個)	
		咳嗽	非咳嗽
実測値 (個)	咳嗽	22	0
	非咳嗽	4	42

表 5 それぞれのアルゴリズムの F 値

モデル	F 値 (%)
Liaqat-based	64.5
Liaqat-based (BPF 未適用)	82.3
Sound Analysis	91.6

た誤検出された音のうち、その多くは会話音であった。この原因として、元々誤検出されやすい会話やくしゃみ等の音が、BPF の適用によってより咳嗽に近い音になってしまったことが考えられる。そのため、今後 BPF を適用する際は、咳嗽の特徴が出やすい範囲以外を全て減衰させるのではなく、対象とする使用環境に応じてフィルタの範囲を調整することを検討している。

また全体の結果より、Sound Analysis の方が検出精度が高かったため、システムを構築する際はこちらを用いる。

6. システムの最大稼働時間評価

6.1 実験内容

患者の音声を常時録音可能な稼働時間が確保されているか検証するために、医師への咳嗽発生状況提示機能を除いたシステムを構築し、実験を行なった。スマートウォッチには Apple Watch Series 4 (watchOS 7.1) を、スマートフォンには iPhone 12 Pro Max (iOS 14.2.1) を用いた。構築したシステムを稼働させ、1 時間ごとのバッテリー残量を計測した。

6.2 結果と考察

実験の結果を表 6 に示す。1 時間ごとの平均バッテリー消費量はスマートウォッチが 19.2%、スマートフォンが 1.73% であった。またシステムの最大稼働時間は 5 時間 12 分であり、現状のシステム構成では患者の音声を常時録音するには不十分であることが分かった。

表 6 各時間におけるスマートウォッチおよびスマートフォンのバッテリー状況

時間	Apple Watch Series 4		iPhone 12 Pro Max	
	残量 (%)	消費量 (%)	残量 (%)	消費量 (%)
1:00	83	17	100	0
2:00	65	18	100	0
3:00	46	19	97	3
4:00	24	22	95	2
5:00	7	17	92	3
5:12	0	7	91	1

システムの最大稼働時間を延長する方法として、録音時の音質を低下させることが有効であると考えられる。具体的にはサンプリング周波数とビット深度を下げることで、および音声フォーマットを圧縮形式にすることで。しかし、同時に咳嗽の検出精度が下がってしまう可能性があるため、これらを変更した場合にバッテリー消費量と咳嗽の検出精度がどのように変化するか調査する必要がある。

7. おわりに

本稿では患者への負担が少ない咳嗽発生時刻の記録方法として、装着時の違和感が少ないスマートウォッチを用いて咳嗽発生時刻を自動的に記録するシステムを提案した。実際の患者の咳嗽音を用いて咳嗽検出アルゴリズムを評価した結果、Liaqat-based よりも Sound Analysis の方が検出精度が高いことが分かった。また実験用のシステムを構築して最大稼働時間を評価した結果、患者の音声を常時録音するには不十分であることが分かった。今後はシステムの最大稼働時間延長に取り組む。十分な稼働時間が確保でき次第、既存の咳嗽発生時刻記録方法に比べて負担が少ないか検証するための実験を行う予定である。

参考文献

- [1] Liaqat, D., Wu, R., Gershon, A. et al.: Challenges with real-world smartwatch based audio monitoring, *Proc. 4th ACM Workshop on Wearable Systems and Applications*, pp. 54–59 (2018).
- [2] Apple: Sound Analysis — Apple Developer Documentation, Apple Developer (online), available from (<https://developer.apple.com/documentation/soundanalysis>) (accessed 2020-11-13).
- [3] Apple: Create ML - 日本語ドキュメント, Apple Developer (オンライン), 入手先 (<https://developer.apple.com/jp/documentation/create-ml/>) (参照 2020-11-13).
- [4] Korpas, J., Vrabc, M., Sadlonova, J. et al.: Analysis of the cough sound frequency in adults and children with bronchial asthma, *Acta Physiologica Hungarica*, Vol. 90, No. 1, pp. 27–34 (2003).
- [5] Freesound: Freesound, Freesound (online), available from (<https://freesound.org/>) (accessed 2020-11-13).
- [6] Hershey, S., Chaudhuri, S., Ellis, D. P. et al.: CNN Architectures for Large-Scale Audio Classification, *Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 131–135 (2017).