

咳音に対しての擬似音節化モデルと軽装着マイクを用いた 咳嗽計数手法の検討

竹之内 翔太郎¹ 沼尾 雅之¹

概要：現在，日本は少子高齢化問題の影響で，要介護者の人数が増え続けているという問題があり，介護者のストレスが原因で虐待事件などが発生している．介護者のストレスの素になる，健康管理に対する負担削減を目的に，本研究では健康状態把握に重要となる咳音を認識する手法の提案を行う．咳音を持つ音響的特性および生体的特徴に着目して，咳音に対して，複数の擬似的な音素に分けて定義を行い，各音素毎に HMM (Hidden Markov Model) を生成して認識を行う．音声の録音には，首にかけることで常時装着した状態での録音が可能な形式のマイクを使用し，特徴量には MFCC (Mel-Frequency Cepstrum Coefficients) に対数エネルギー項 *LogPow* を加えたものを使用した．まず，簡易的な声道モデルを作成してマイクの仕様を決定し，続いて 3 名の被験者から，複数回の咳嗽からの計数を行う実験と発話中に咳嗽を挟んだ音声を認識させる実験を行った．実験結果から，1 回の咳音に対しスペクトログラム情報より 2 分割して音節化し，「咳音の直前に無音状態が入る」という認識時の文法ルールを定義して認識を行った場合，発話せずに複数回咳を行ったときに 95% 以上の適合率 (Precision) が得られ，咳嗽 10 回につき平均 9 回以上検出されるという結果となった．また，発話中に咳嗽を含んだ音声を認識させる実験でも，81% 以上の適合率となった．

Study of cough-counting method by using pseudo-syllable model for cough-sound and light-wear type microphone

SHOTARO TAKENOUCHI¹ MASAYUKI NUMAO¹

1. はじめに

現在，我が国日本では少子高齢化問題が深刻なものとなっている．総務省統計局の統計データによれば，高齢化率 (65 歳以上の高齢者が総人口に占める割合) は 2015 年時点で 26.7% を占めており，2060 年には 40% 近くを占めると予測されている．これにより，介護を要する高齢者の数が増える一方で，介護者自身の高齢化も進むことが予想されるため，介護者のストレス悪化が原因となる虐待事件がさらに多くなる危険がある．このように急速に進む少子高齢化社会に対応するために，介護に対するストレスの削減を目的とした，見守りの自動化への貢献を図る研究が近年行われている．見守りの自動化を行うことができれば，介

護のストレスは緩和され，介護者と高齢者が両方ともに生活がより快適になることが期待できる．

本研究では，人から出される咳嗽を認識しカウントすることで，見守り自動化への寄与を狙う．咳嗽の回数や傾向を把握することで健康状態の把握を行うことで，病気の発見に繋がる可能性がある．また，認識された咳嗽の部分を抽出して，医者に送ることにより遠隔診断への貢献も期待できる．咳音の認識を行うために，通常の話声での音声認識で使われている HMM (Hidden Markov Model) を用いて，認識を行う．HMM で話声を認識する場合，一般的に音素に分けて HMM を定義し，各音素の HMM を連結させる形で単語の HMM を定義している．例えば，単語「感謝」に対する HMM は /k/, /a/, /N/, /sh/, /a/ の計 5 つの HMM が連結される形で定義している．これと同様に，咳音の認識に用いる HMM についても複数の HMM を連結させる形式で定義を行う．ただし，咳音は発話音とは違い，言葉で

¹ 電気通信大学大学院 情報理工研究科
情報・ネットワーク工学専攻
The University of Electro-Communications

はない「非言語音声」にあたるため、上記のように一般的な音素に分けることは困難である。そこで咳音に対して、音響的特徴だけでなく生体的特徴について考察を行い、その上で咳音に対する擬似的な音素に分けて HMM を生成して認識を行い、どういう風に音節化すれば正確に咳嗽を計数できるか考えていくこととする。また、音声の採集にあたって、装着型マイクを用いて録音を行う。

2. 関連技術

2.1 ロボットを用いた見守りシステムに関する研究

住谷らは [3], 移動ロボット Kobuki を用いた見守りシステムを提案した。Kobuki と Kinect を PC に接続し、Kinect からは対象者の距離や関節位置の情報から、転倒の判定を行うようにしている。Kobuki は Kinect からの情報に従って対象者を追うという方式となっている。Kinect から取得した骨格情報のうち、頭部と両膝の Y 座標の差が閾値未満となった状態が 5 秒以上続いた場合、対象者が転倒したとみなし、予め指定されたアドレスにメールを送って通報するという仕組みになっている。

2.2 音情報を用いた見守りシステムに関する研究

西らは [4], 高齢者の生命的危機を検知するために、音情報を利用し、人が倒れたときに出る音のような、日常では現れないような音（生活異常音）、あるいは日常的に生じる生活音が一定時間観測されなかった場合に連絡先に通報するシステムを提案した。生活音と異常音を識別するために、連続 DP マッチング法に加えて新たに考案された、距離マトリクス画像からの直線検出によるワードスポッティング法が適用されている。

2.3 非言語音声認識に関する研究

主に見守りや健康状態管理を目的に、咳嗽のような非言語音声を認識対象とした研究はいくつか行われている。

山田らは [5], 個人差を少なくして、嚥下音から水分摂取量を自動的に推定する手法について研究を行った。嚥下音、発話、深呼吸、咀嚼音、咳の 5 種類の音から、HMM で嚥下音を分類して、特徴量として、パラ言語情報に向いている音響的特徴量セットである LLD (Lower Level Description) 特徴量セットから嚥下音に相関性が高いと考えられる MFCC (Mel-Frequency Cepstrum Coefficients) 及び継続時間 (Duration) を用いて、より高精度で個人差の少ない水分摂取量推定を図った。結果、1 日に必要な水分摂取量 1,500ml から 500ml の誤差まで精度を向上できた。

咳嗽認識を主目的とした研究として、道幸らは [6] 誤嚥した際に出る咳嗽反射の検出を目的に、スロートマイク (咽喉マイクロフォン) からの口腔咽喉音より、波形処理 (絶対値処理と振幅包絡線処理) を行ってから、包絡線の情報 (振幅の最大値など) を用いて咳嗽を検出するシステムを提案し

た。また Zhu らは [7], DTW (Dynamic Time Wrapping) を用いた咳嗽認識手法を提案した。特徴量として、0~12 次元の MFCC の各要素に 1 次差分を加えた、計 26 次元の特徴ベクトルを用いている。34~53 歳の男女 3 人の被験者からの咳嗽を認識を行った結果、咳嗽認識率 (本来咳である音が、咳として正しく認識された割合) が、被験者ごとに最高 94.2% 最低 86.2% の数値を達成した。

3. 提案手法

3.1 従来手法の課題点

[3] で使われた、移動ロボット Kobuki の値段は、[8] によれば約 7 万 8 千円とかなり高いものになっているので、見守りを行うシステムを広く普及させたい場合、導入費用がなるべくかからず手軽なものにする方が望ましい。また、[5] では、咳音などに対する HMM の具体的な状態定義に触れられてなく、[6] では認識対象が咳音のみであり、システムの実用を考えると発話との識別検証がなされる必要がある。そのため、咳音を持つ様々な特質を利用して音響モデル定義を行い、その上で話声を認識対象に加えて咳音の認識を行う必要があると考えた。

3.2 提案手法の概要

咳音に対する音響的特徴などの情報を利用して、咳音を持つ他種の音に無いような特徴について考察し、1 つの咳音の HMM を擬似的な音節を持つ HMM の連続として定義し、各音節に対する正解区間のラベリングをして特徴量に基づいて教師あり学習をして HMM の学習を行う。HMM の学習後は、認識に関する文法ルールを定義してそれに基づき認識を行う。HMM の学習から、認識結果が導出されるまでの大まかな流れは、以下の図 1 のようになっている。

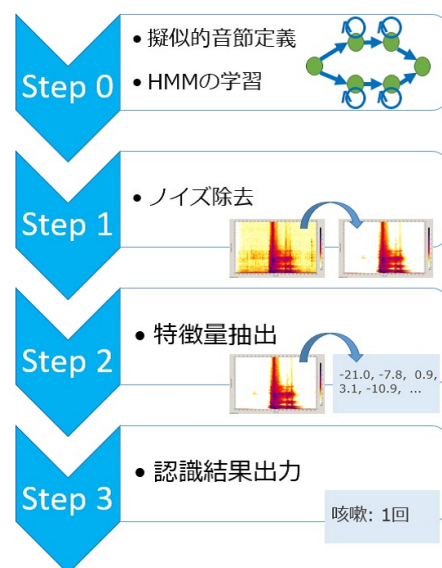


図 1 認識までの一連の流れ

特徴量にエネルギー項をさらに加えたのは、図3において咳の前半部の周波数成分強度が全体的に強いので、他の音との区別ができる特徴になることを期待したためである。

3.3 予備実験の概要

3.3.1 実験1:マイクの概要と仕様選定

まず、正確な認識にあたり、最も適切なチューブ型マイクのチューブ長の選定を行う。認識にあたり、録音に用いたマイクは、下の図5のようにチューブ型マイクを用いて録音している。なお、発生者間での首囲の差による録音音声の影響を極力避けるため、チューブを喉（首の前部分）に直接当てる形で録音を行っている。



図5 チューブ型マイクの着用

このマイクに音声が入る仕組みは、チューブ内の空気振動に関する情報によるものである。チューブ内の空気は、気柱に該当するものと考えられるため、チューブの長さによっては共鳴が発生し大きな音が取れる可能性が期待できる。そこで、チューブの長さを変えて録音を行い、最も良い精度となったチューブの長さを採用することとする。

音声の質のばらつきによって、認識結果に影響を及ぼすことを防ぐため、図6のように人間の声道を簡易的に模したモデルを作成して行う。このモデルは、長方形のボトルに円形のケースが内接する形で入っており、ボトルとケースの間に水を満たすことにより、人間の声道と首との関係を簡易的に再現している。また、円形のケースの上端部を開口することで、声が口から出されることの再現を行っている。なお、モデルの周囲長 (Perimeter) は36cmであり、一般的な男性の首囲に近いものとなっている。



図6 簡易的声道モデル

また、真上から見た、人間の首部と声帯との位置関係を簡易的声帯モデルと比較した図は、図7のようになっている。

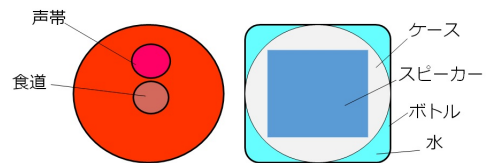


図7 真上から見た首部（左）と簡易的声帯モデル（右）との比較

このモデルの中にスピーカー（ボトル内にある水色のブロック状の物体）を入れてそこから音声を再生することで、チューブの長さを変えても同じ原音を再生できるようにしている。なお、前記と同様の理由により、モデルの周囲長の影響を抑えるため、スピーカーから音が出る場所にマイクを極力近づける形で録音を行った。

3.3.2 実験2:咳音認識に適した擬似的音節の検討

チューブの長さを決定したら、複数人の被験者からの咳音認識を行う。咳音の性質から、候補とした擬似的音節化モデルと、それに対するHMMと文法ルールに従い、認識を行い咳嗽のカウント精度についての検討を行う。実用性を考慮し、以下のような2通りの方式の音声を認識させることにする。

- 連続で複数回咳をしたもの
- 話声の中に、数カ所で咳を挟んだもの

咳音に対するHMMの学習時には、学習用の音声データとして合計25回分の咳嗽を行った。その際、咳には多くのパターンが存在していることから、咳の音量に大小を付けたり連続で2回咳をしたりするなどして、バリエーションを多くして行うようにした。なお、不特定話者に対する認識を想定し、学習用に録音した音声の発声者は1名だけにし、学習時に用いた音声とHMMは、全被験者で統一して同じものを使用することにした。

4. 評価実験

4.1 実験1の結果

最適なチューブ型マイクの長さを決めるため、チューブの長さを46~54cmの間で2cm刻みで行った。なお、認識には音節定義Iを用いている。認識に用いた音声データは、発話音と咳音（1回分）それぞれ30個ずつである。

認識を行った結果、チューブの長さとのF値の関係は図8のようになった。なお、F値 (F-measure) は適合率 (Precision) と再現率 (Recall) の調和平均であり、下記のようにして求めた。

$$F\text{-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

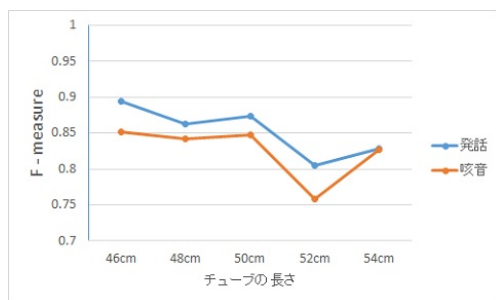


図 8 チューブの長さに対する認識結果の F 値

咳音に対する F-値の結果が最良であり、発話音に対しても最良であったため、マイクのチューブ長を 46cm にして使用することとした。この長さでの認識結果が良かった理由として、咳音がチューブ内で共鳴が起きて音量の大きい音が出たためと考えられ、原音が同じ咳音をスピーカーから再生してパワーの最大値を計測した結果、チューブ長が 46cm（認識結果が最良）の時に 36.34dB、54cm（認識結果が最悪）の時に 28.61dB となり、7.73dB のパワー差があった。また、咳音に対する周波数スペクトラムを求めたところ、図 9 のようになり 525Hz の成分の強さが最も強くなることが分かり、この周波数成分が咳音の固有周波数との関連性があると考えられた。

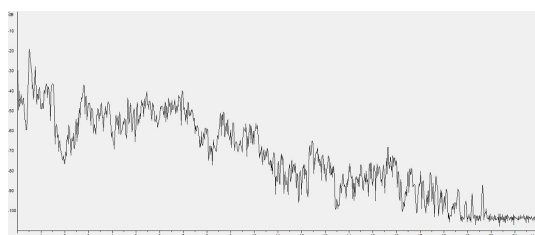


図 9 咳音に対する周波数スペクトラム

4.2 実験 2 の結果

実験結果で発声者間で個人差が大きくないかを検証するため、被験者 A, B, C の計 3 名を取って実験を行った。なお、被験者は全員とも 20 代男性である。

咳を 10 回分行い、認識された咳の回数の計測を行う実験を実験 2.1、話している途中で咳をしたものを認識させる実験を実験 2.2 とする。そして、咳音に対して定義した擬似的音節毎に精度の評価を行う。具体的には、最初は咳と認識された時刻区間で本当に正しい区間であったものの割合、すなわち適合率の測定を行い、それから認識された咳の回数の平均および分散の評価を行う。

4.2.1 実験 2.1 の結果

まず、咳嗽に対する適合率は図 10 のようになった。なお、被験者 A は計 7 セット分、被験者 B と C は計 5 セット分の音声を認識に使用している。音節定義 I を用いた場合、適合率は全被験者で 95% を超える結果になった。

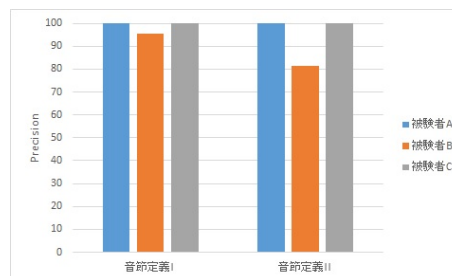


図 10 実験 2.1 における咳の適合率

続いて、認識された咳の回数の平均と分散を求めた結果、図 11 と図 12 のようになった。

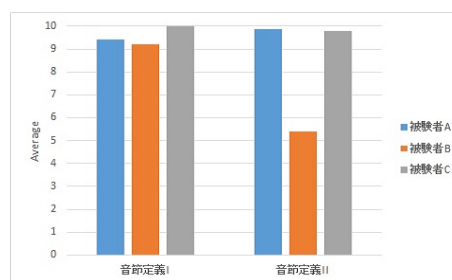


図 11 実験 2.1 で認識された咳の回数の平均

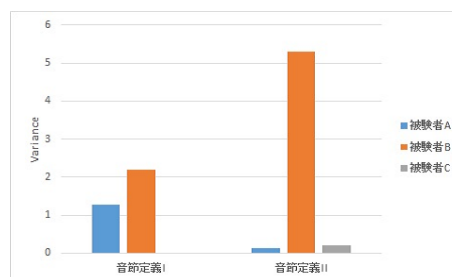


図 12 実験 2.1 で認識された咳の回数の分散

これらの結果から、何の発話も行わず咳のみを行った場合、音節定義 I のほうが全体的に正確でかつ個人差による結果のばらつきが少なくなるという結果になった。

4.2.2 実験 2.2 の結果

実験 2.2 では、計 5 種類のテキストの中に、咳を 3 ケ所で行った音声から咳嗽を認識するという実験を行った。

まず、咳嗽に対する適合率は図 13 のようになった。音節定義 I を用いた場合、適合率は全被験者で 81% 以上という結果になった。

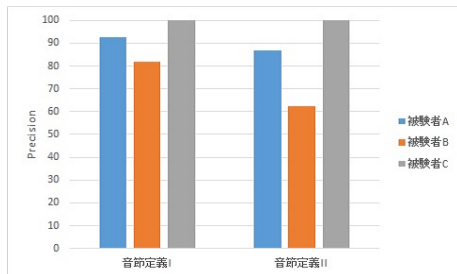


図 13 実験 2.2 における咳の適合率

続いて、認識された咳の回数の平均と分散を求めた結果、図 14 と図 15 のようになった。

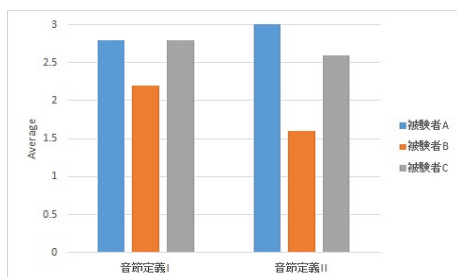


図 14 実験 2.2 で認識された咳の回数の平均

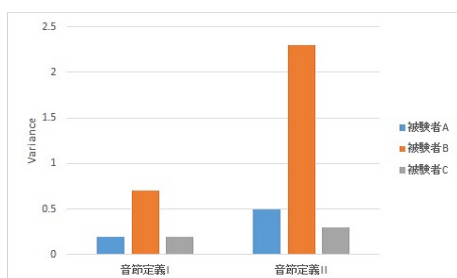


図 15 実験 2.2 で認識された咳の回数の分散

実験結果から、実験 2.1 の時と同様に音節定義 I のほうが全体的に正確かつ個人差による結果のばらつきが少なくなるという結果になった。音節定義 II を用いた場合の結果は全般的に悪く、特に被験者 B に対する結果は、前の実験時よりも悪化したものとなった。

5. おわりに

本研究では、健康状態情報のサインとなる側面を持つ咳嗽を正確に認識してカウントするために、咳嗽がもつ音響的特性などを分析し、咳音に対して擬似的な音節定義を行う手法の提案を行った。

咳音を前半と後半の 2 つに分けて音節化し、それぞれに対する HMM を学習して生成することで、適合率に関して良い結果が得られ、平均認識回数に関しても正解の回数に近く、良い結果が得られた。特に、咳のみを行った場合の適合率は全被験者とも 95% を超える結果となった。しかし、認識対象の音声を発声した計 3 人の被験者は、いずれとも同じ性別であり年齢もほとんど近いいため、年齢差などが原因で咳嗽の認識精度に大きな個人差が出る可能性が存在する。本研究の目的は、介護での見守り自動化への貢献であるため、実用を考えると、高齢者に対して同様の実験を行い検証する必要がある。

また、本研究で特徴量の一部に取り入れた対数エネルギー項は、咳嗽のように音量が大きな音声の識別が容易になるものの、音量が小さな音の場合は無音状態との区別が難しいため、精度が落ちやすい。そのため、健康管理や見守りに役立つ咳嗽以外の非言語音声、すなわち嚙下音や通常の呼吸音などの認識は精度が悪くなりやすいので、咳音に加えて他の非言語音声を認識対象とした時はエネルギー項以外の特徴量を考えていく必要性が出て来る可能性が存在する。それに加えて、今回提案した簡易的声帯モデルでは、ボトルの直径が 8.5cm 以上であったが、人間の声帯がある気管の内径が成人の場合 1.5cm 程度であり、実際の値と大きな差が存在する。そのため、より最適なマイクのチューブ長を探すためには、本研究で提案した簡易的声道モデルを、より実際の人間の喉部と声帯との関係に近いものに改良していく必要がある。

謝辞

本研究は、JSPS 科研費 JP17H01823 「無負荷センサ統合による見守りシステムの構築法」の助成を受けたものです。

参考文献

- [1] 総務省統計局: 統計トピックス No.90 統計からみた我が国の高齢者 (65 歳以上) — 「敬老の日」にちなんで —, 2015
- [2] 日本気管食道科学会, 「気管食道科に関連する疾患・症状-せき」, 入手先 (<http://www.kishoku.gr.jp/public/disease07.html>) (参照 2017-05-05).
- [3] 住谷拓馬, 菅谷みどり, “家庭用移動ロボットを用いた見守りシステムの提案”, 情報処理学会研究報告, Vol.2014-SE-184 No.5, Vol.2014-EMB-33 No.5, 2014/5/19, pp.1-7.
- [4] 西宏之, 金金, 木村義政, 柿木稔男, “音情報を用いたお年寄り見守りシステムの提案”, 電子情報通信学会, 信学技法, ICM2013-49, LOIS2013-53(2014-01), pp.81-84.
- [5] 山田侑太郎, 西村雅史, 峰野博史, “個人差の無い特徴

- 量を用いた水分摂取量推定手法の検討”, DICOMO2016, pp.1814-1820 (2016).
- [6] 道幸成久, 辻村肇, 松村雅史, ”咽喉マイクロフォンによる高齢者の咳嗽の無拘束モニタリング～MSP430 搭載カードサイズデータ収集システム～”, 電子情報通信学会, 信学技法, MBE2012-18(2012-6), pp.27-30.
 - [7] Zhu, Chunmei, Baojun Liu, and Ping Li. ”Cough Recognition Based on Mel Frequency Cepstral Coefficients and Dynamic Time Warping.” Mechanical Engineering and Control Systems: Proceedings of the 2015 International Conference on Mechanical Engineering and Control Systems (MECS2015). World Scientific, 2016.
 - [8] Kobuki | 研究用移動ロボット「コブキ」, 入手先 <<https://www.rakunew.com/items/71714>>, (参照 2017-05-08)
 - [9] SoX - Sound eXchange, 入手先 <<http://sox.sourceforge.net/>>, (参照 2017-01-17)
 - [10] 情報処理学会: 情報処理学会論文誌 (IPSJ Journal) 原稿執筆案内, 情報処理学会 (オンライン), 入手先 <http://www.ipsj.or.jp/journal/submit/ronbun_j_prms.html> (参照 2017-03-01).
 - [11] 大正富山医薬品株式会社, ”咳音を, 聴く”, 入手先 <<http://クラリスweb.com/listening/pertussis.html>>, (参照 2017-05-07)
 - [12] 第一三共ヘルスケア, 咳 (せき) の原因—くすりと健康の情報局, 入手先 <http://www.daiichisankyo-hc.co.jp/health/symptom/20_seki/index1.html>, (参照 2017-05-07).
 - [13] 白クマ先生の子ども診療所, 「せき, 喘息」, 入手先 <https://www.med.or.jp/clinic/sick_seki.html>, (参照 2017-05-07).