

呼吸音検出アルゴリズムの為の音響特徴量の抽出*

○和井田将広, 加科優希, 中島弘史, 三好和憲 (工学院大)

1 はじめに

日中の眠気や怠倦感を引き起こす原因の一つとして睡眠時無呼吸症候群 (sleep apnea syndrome 以後 SAS と称す) がある. SAS とは睡眠中に無呼吸 (10 秒以上の呼吸停止) 状態が 1 時間当たり 5 回以上繰り返される病気である^[1]. SAS 患者の呼吸の特徴として, 激しいいびきの後しばらく呼吸が止まり, その後あえぐような激しい息またはいびきで呼吸が再開するといったものがある^[2].

SAS の診断方法で代表的なものとして入院検査 (睡眠ポリグラフ検査) とパルスオキシメーターを使用した方法がある. 入院検査は正確に診断ができるが, 入院が必要なため高額であり, センサーを多く付けているため普段の睡眠状態とは大きく異なる状態になるというデメリットがある. パルスオキシメーターは自宅でもできるというメリットがあるが, 体にセンサーを付けて行うため, 普段の睡眠状態と完全には一致しない.

最近では SAS の診断を目的とした, スマートフォン用アプリである「いびきレコーダ pro」^[3]などがリリースされている. このアプリはいびき音の録音や無呼吸状態になる箇所を認識し知らせるなどの機能がある. しかし無呼吸状態を認識する機能は有料版のみにあり, 無料版ではその機能がなくコストがかかる. また詳細なアルゴリズムが非公開であるため検出精度などが不明確である. このため従来よりも安価で測定の精度の高い簡易診断システムの構築が望まれる.

2 本研究の目的, 長期計画, アプローチ

2.1 目的

低コストで精度の高い呼吸音による睡眠時無呼吸症候群の簡易診断システムの構築を目的とする.

2.2 本研究の長期計画

本研究「睡眠時無呼吸症候群の簡易診断システムの構築」は JSPS 科研費 24500213 の

助成を受けたものである. 四カ年計画 (Table 1) の研究の一部である. 本年度は 2 年目にあたり, 呼吸音検出システムの構築のために (C) 「呼吸音の特徴量抽出」と (D) 「判別器の作成」を行った. また, 近年公開され実用化された製品である睡眠アプリや睡眠計との比較をするために (E) 「睡眠時環境音の SN 比評価及び睡眠アプリ・睡眠計の調査」を行った. 本研究は (C) について報告する.

Table 1 本研究の四カ年計画について

予備検討期間 (2011 年度)	呼吸音による睡眠状態判別ソフトウェアの構築 睡眠時の呼吸音検出に関する研究
1 年目 (2012 年度)	(A) 睡眠時における音環境調査 (B) ハードウェア要件の明確化
2 年目 (2013 年度)	(C) 呼吸音の特徴量抽出 (D) 判別器の作成及び評価 (E) 睡眠時環境音の SN 比評価及び睡眠アプリ・睡眠計の調査
3 年目 (2014 年度)	精度評価及びパラメータチューニングによる精度向上
4 年目 (2015 年度)	SAS の簡易診断システムを構築

2.3 アプローチ

目的を達成するため, 複数人の睡眠時の呼吸音を収録し, SAS の検知に重要であると考えられるいびきの特徴量を抽出する^[4].

3 特徴量の抽出

本研究では, 特徴量としてピーク周波数, エンベロープ, メル周波数ケプストラム係数 (Mel Frequency Cepstral Coefficient 以降 MFCC と称す), フォルマント周波数を抽出した. 特徴量抽出には, 信号処理やグラフィック表示に長けている MathWorks 社が開発した言語 MATLAB を使用した.

3.1 前処理

特徴量抽出の前処理として, 音源 $y(t)$ をフレーム長 f_s , シフト長 c_s で分割し, ハン窓

* Sounds feature extraction for breath state detection algorithm, by WAIDA, Masahiro and KASINA, Yuuki and NAKAJIMA, Hirofumi and MIYOSI, Kazunori (Kogakuin University).

をかけた。ただし、 t は時間である。

3.2 ピーク周波数

数納氏と神藤氏による研究^[5]からいびきの主な周波数は 100~500 Hz であることが明らかとなっているので、ピーク周波数が特徴量として有効であると推定できる。ピーク周波数の計算の概要を以下に示す。

音源 $y(t)$ をフーリエ変換 (FFT) し、 $Y(\omega)$ を求めた。ただし ω は角周波数である。次式で周波数毎のパワースペクトルレベル $L_p(\omega)$ を次式で求める。

$$L_p(\omega) = 10 \log_{10}(|Y(\omega)|^2) \quad (1)$$

$L_p(\omega)$ から最大値 L_{\max} とその周波数 f_{\max} を求める (式 3~4)。

$$L_{\max} = \text{Max}\{L_p(\omega)\} \quad (2)$$

$$f_{\max} = \text{Argmax}\{L_p(\omega)\} \quad (3)$$

ただし $\text{Max}\{f(x)\}$ は $f(x)$ の最大値を求める関数であり、 $\text{Argmax}\{f(x)\}$ は $f(x)$ が最大となる x を与える関数である。

3.3 エンベロープ

いびきは周期的に振幅が変化するので、振幅の概形を抽出したエンベロープが特徴量として有効であると考えられる。エンベロープの抽出法を以下に示す。音源 $y(t)$ から負の周波数成分を 0 とし、正の周波数成分だけを持つ解析信号を算出する^[6]。また、全周波数帯域でエンベロープを抽出すると主に高周波のノイズにより、なめらかな概形が得られないため、 $\omega_1 \sim \omega_2$ の角周波数成分を抽出した (次式)。

$$Y(\omega) \begin{cases} 0 & (\omega < \omega_1 \text{ or } \omega > \omega_2) \\ 2Y(\omega) & (\omega_1 \leq \omega \leq \omega_2) \end{cases} \quad (4)$$

今回は主に電源ノイズを除去するため下限は 60 Hz、また上限はなめらかな概形が得られる値を目視で確認したところ 110 Hz 付近が最適であったため、 $\omega_1 = 2\pi \cdot 60$ 、 $\omega_2 = 2\pi \cdot 110$ とした。式(5)で算出した $Y(\omega)$ を逆フーリエ変換 (IFFT) する。これで解析信号 $s(t)$ が求められる。 $s(t)$ の絶対値をとることで、エンベロープ波形 $e(t)$ が求められる。

$$e(t) = |s(t)| \quad (5)$$

3.4 MFCC^[7]

MFCC とは、周波数スペクトルを対数ケプストラムに変換し、低次成分に対して人の周波数知覚特性を考慮した重み付けをした特徴量である。MFCC の計算の概要を以下に示

す。音源 $y(t)$ をフーリエ変換した音声波形 $Y(\omega)$ の絶対値を算出し、振幅スペクトル $S(\omega)$ を次式で求める。

$$S(\omega) = |Y(\omega)| \quad (6)$$

$S(\omega)$ にメルフィルタバンクをかける。メルフィルタバンクとはメル尺度上で等間隔な三角窓を複数合わせたフィルタである。これによりそれぞれの三角窓に対応した周波数成分が強調される。今回は 20 次元の MFCC を用いた。強調された周波数帯域毎の $S(\omega)$ の和を求め、周波数帯域毎に 1 つの値を出す。振幅スペクトルの対数を取り、コサイン変換 (DCT) をする。低次成分が声道特性の成分、高次成分が音源成分に分離されたケプストラムが求められる^[8]。声道特性の成分が重要と考えられるため、低次成分のみを取り出す。今回は 12 次元目以下を低次成分とした。

3.5 フォルマント周波数^[9]

フォルマント周波数とは、音声のスペクトル包絡上で特定のピークであり、第一ホルマント周波数は閉鎖型無呼吸症候群 (OSA) の診断において高い診断能を示すという報告がある^[4]。そのため特徴量として有効であると考えられる。以下にフォルマント周波数の抽出法を示す。今回は、過去の信号から未知の信号を予測する LPC 分析^[9]を用いて、スペクトル包絡を求めた。LPC 次数を k とし、音源 $y(t)$ の自己相関関数 $u(i)$ を次式で求める。

$$u(i) = \frac{1}{N} \sum_{t=i}^{N-i} y(t)y(t-i) \quad (i=0,1,\dots,k) \quad (7)$$

ただし N はデータの総数である。

$u(i)$ からユール・ウォーカー方程式をレビンソン・ダービン再起法を用いて解き、LPC 係数 $a(i)$ を $i=1,2,\dots,k$ について求める^[9]。初期値として残差パワー $v(0)=u(0)$ とする。LPC 係数の周波数応答の絶対値より振幅スペクトル $S(\omega)$ を求め、元の波形の振幅スペクトルの包絡を得る。この包絡の頂点かつ最も周波数の低い点を第一フォルマント周波数 F とする。

4 特徴量の評価実験

実際に収録した睡眠時の音声データのピーク周波数、エンベロープ、MFCC、フォルマント周波数の特徴量の評価を行う。なお、本研究では波形から読み取れる主観的な評価を行っており、機械学習を用いた定量的な評価

は加科氏による研究「呼吸音検出アルゴリズムにおける判別器の作成と評価」^[10]で報告する。音声データは1年目の研究^[5]で収録したデータを使用している。

4.1 使用機材について

機材はFig. 1の(a)精密騒音計(ONOSOKKI LA-4440), (b)普通騒音計(RION NA-20), (c)ICレコーダー(TASCAM DR-05), (d)校正器(Brüel & Kjær TYPE4231)を使用した。音声データには精密騒音計で収録した音と普通騒音計で収録した音の2種類があり, 今回は精密騒音計で収録した音声データを使用した。



Fig. 1 使用機材

実際に収録した音声のスペクトログラムをFig. 2に示す。

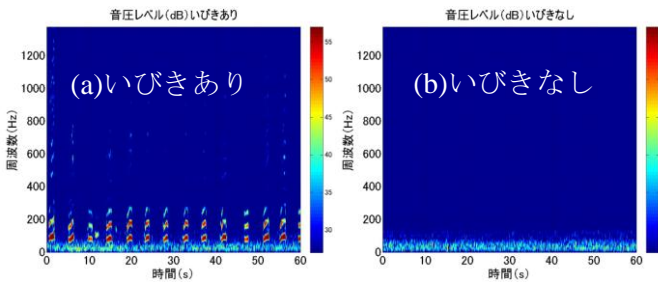


Fig. 2 収録した音声のスペクトログラム

4.2 ピーク周波数

フレーム長 $f=0.5$ s の時の時間毎のピーク周波数の変化の波形をFig. 3に示す。

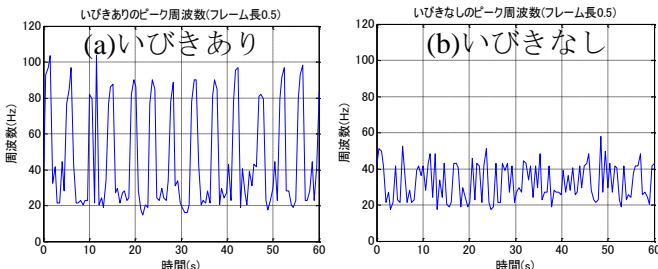


Fig. 3 時間毎のピーク周波数の変化

今回抽出されたピーク周波数は約 70~120 Hz の範囲にいびき音が出ており, 70 Hz より下の周波数のほとんどはノイズであった。また, いびきの部分は時間毎の変動が大きく, ノイズの部分は時間毎の変動が小さかった。このことから, ピーク周波数はいびきの判別

に有効であると考えられる。

4.3 エンベロープ

エンベロープ波形をにFig. 4示す。

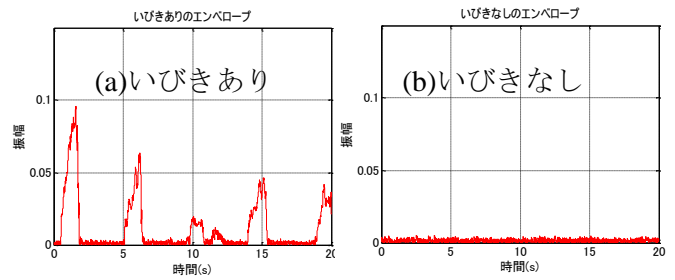


Fig. 4 エンベロープ波形

いびきの部分は振幅が大きくなり, いびきなしの振幅はほぼ低い値で一定になっていた。また, いびきなしからいびきありへの変化がはっきりしているため, エンベロープ波形はいびきの判別に有効であると考えられる。

4.4 MFCC

1 分間の音声波形を短いフレーム長で切り出し MFCC を求めた時の 20 次元のスペクトログラムをFig. 5に示す。

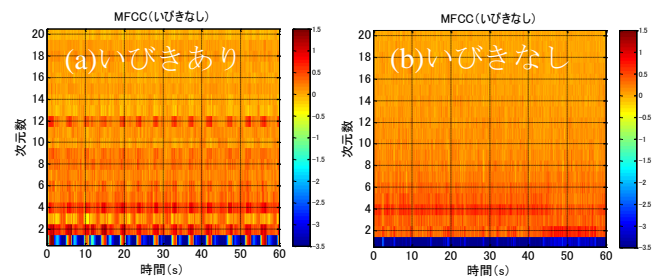


Fig. 5 MFCC

高次成分の違いは目視では判別できないが, 低次成分はわずかに違いが見られる。特に 1 次元目は, いびきがある時間帯の値が少し大きくなっている。これは MFCC が有音声だと 1 次元目の値が大きくなり, 無音声だと値が小さくなることによるものと思われる^[11]。このことから MFCC はいびきの判別に有効であると考えられる。

4.5 フォルマント周波数

LPC分析によって抽出したスペクトル包絡の波形をFig. 6に示す。

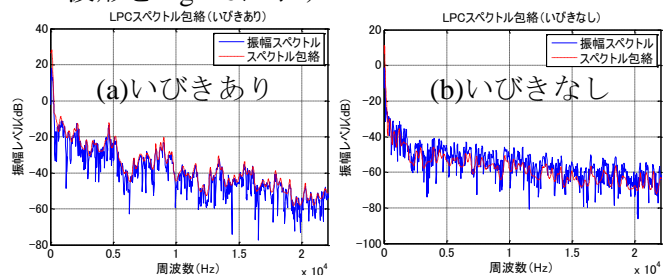


Fig. 6 LPC スペクトル包絡

今回はLPC 次数を 64 としてLPC 分析を行った。いびきありの方では、100 Hz 付近に頂点が見られる。ピーク周波数でも同様の結果が得られているため、正確な値が得られていると考えられる。いびきなしの方では 50 Hz 付近に最大振幅があるが、電源系の低周波ノイズと考えられる。1000 Hz 付近のわずかな山についても収録時の定常的なノイズである可能性が高い。次に時間毎のフォルマント周波数の変化の波形をFig. 7に示す。

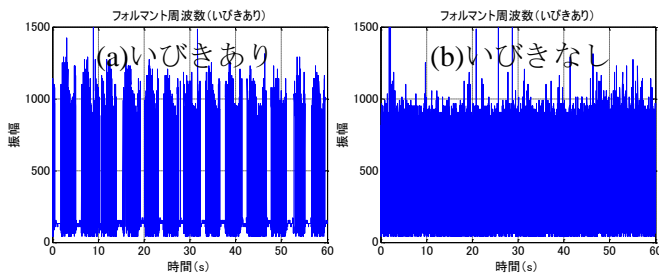


Fig. 7 時間毎のフォルマント周波数の変化
いびき音のある部分は約 100~140 Hz に集中している。いびき音がない部分では約 50~1500 Hz の間で大きく変動をしていた。その部分はほとんど雑音である。いびきの部分については雑音の影響をほとんど受けずに安定した値が得られることから、ピーク周波数よりも雑音に対してロバストである。そのため、フォルマント周波数はいびきの判別に有効である。

5 まとめと今後の課題

ピーク周波数では、いびきのピーク周波数は 70 Hz 以上になっており、それ以下の周波数帯にはほとんど雑音しかないことが分かった。雑音を除去すればいびき音をより明確に抽出できると考えられる。エンベロープではいびきがある時とない時の差をはっきりさせることができた。しかし騒音の多い環境やいびき音が小さい場合でも同様の傾向が表れるのか検証が必要である。MFCC は大きいいびき音についての機械学習による判別結果^[10]は誤分類率がほぼ 0 であった。しかし弱いいびき音は誤分類率は高くなった^[10]。原因としては雑音といびきの波形の形が似ていた、いびき音が小さく認識されない等が考えられる。フォルマント周波数では第一フォルマントからいびきの周波数を抽出しやすかったが、雑音の周波数の変化が大きいため、ノイズリダクションを行うことによって

より分かりやすい特徴量になると考えられる。

より精度を高くするために、新しい特徴量を導入することや、多くのデータをとるために多くの人の睡眠時の呼吸音を収録することが今後の課題である。

6 謝辞

本研究を進めるにあたり、適切なお助言とご指導を頂いた鶴秀生博士に感謝致します。また医学的側面から適切なお助言を賜って下さった東京大学附属病院の山口泰弘先生に感謝致します。

7 参考文献

- [1] 「睡眠時無呼吸症候群サイト」
<http://www.sleep.or.jp/>
- [2] 「みやびクリニック」
<http://www.miyabi-clinic.jp/index.jsp>
- [3] 「いびきレコーダ Pro」Google Play の Android アプリ
<https://play.google.com/store/apps/details?id=com.MusicalSoundLab.IntelligentRecorder>
- [4] 榎本崇宏「鼾解析による新しい閉塞型無呼吸の診断支援に関する研究」科学研究費助成事業(科学研究費補助金)研究成果報告書, 2012.
- [5] 数納聖冴, 神藤徳彦「睡眠時無呼吸症候群の簡易診断システムの構築のための音環境調査とハードウェア要件の明確化」信学技報, vol.112, pp.1-6, 2013.
- [6] 芦野隆一, 萬代武史, 守本晃(「連続ウェーブレット変換について」国際数理科学協会会報, no.72, 2010.
- [7] 森山剛, 茂美穂「音声感情を含むことによる声室変化の解析」東京工芸大学工学部紀要, vol.34, no.1, pp.58-64, 2011.
- [8] 古井貞熙著「新音響・音声工学」近代科学社, 2006.
- [9] 日本音響学会「新版音響用語辞典」コロナ社, 2003.
- [10] 加科優希「呼吸音検出アルゴリズムにおける判別器の作成と評価」音講論集 2014 春, 2-P4-23, 2014
- [11] 金澤常助「有音声/無音性情報を用いた単語認識」日本大学理工学部学術講演会論文集, pp.495-496, 2010.