

呼吸音検出システム構築のための特徴量の抽出 Sounds feature extraction for breath state detection system

和井田 将広 (コンピュータ科学科)

Masahiro Waida

数理音響学研究室

指導教員

中島弘史 准教授

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. 本研究の目的と長期計画及びアプローチ

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。四ヵ年計画(表 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、公開された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。

表 1 診断システム構築の四ヵ年計画について

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) 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 ピーク周波数 ピーク周波数とは、波形に含まれる周波数成分の中で最も強い周波数のことである。先行研究の「睡眠時の呼吸音検出に関する研究」[5]の成果からいびきの主な周波数は 85~200Hz 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。

元の音源を $y(t)$ とする。ただし t は時間である。今回は切り出す長さ (フレーム長) を 0.5[s] とし、切り出した波形にブラックマン窓を掛けた。それをフーリエ変換 (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を表したパワースペクトル $P(\omega)$ を求める。

$$P(\omega) = |Y(\omega)|^2 \quad (1)$$

P の常用対数を出し 10 を掛け、周波数毎の強さをレベルで表した音圧レベル $L_p(\omega)$ を求める。

$$L_p(\omega) = 10 \log_{10}(P(\omega)) \quad (2)$$

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

$$L_{\max} = M\{L_p(\omega)\} \quad (3)$$

$$f_{\max} = \text{Argmax}\{L_p(\omega)\} \quad (4)$$

ただし $M\{f(x)\}$ 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。

これでピーク周波数が求められる。実際には時間毎の変化を見るため短いフレームで切り出した波形毎にピーク周波数を求める。

3.2 エンベロープ エンベロープとは、波形の上部の概形をとることである。波形の変化が分かりやすくなるので、通常の波形とは違った特徴量が抽出できると考えられる。エンベロープ抽出法を以下に示す。

音声波形 $y(t)$ から負の周波数成分を無くし、正の周波数成分だけを持つ解析信号を算出する。フーリエ変換した波形 $Y(\omega)$ の負の周波数成分とは、FFT による処理の場合ナイキスト周波数から最大周波数までの周波数成分に相当する。また、そのままエンベロープを抽出しただけでは高周波成分や低周波のノイズにより、綺麗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)。

$$Y(\omega) \begin{cases} 0 & (\omega < \omega_1 \text{ or } \omega > \omega_2) \\ 2Y(\omega) & (\omega_1 \leq \omega \leq \omega_2) \end{cases} \quad (5)$$

今回は $\omega_1 = 2\pi \cdot 60$, $\omega_2 = 2\pi \cdot 110$ とした。

式(5)で算出した $Y(\omega)$ を時間領域の波形に変換する為に逆フーリエ変換 (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。

$$e(t) = |s(t)| \quad (6)$$

3.3 MFCC[6] 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。声道特性のある低次成分が引き伸ばされるため、いびきの特徴が分かりや

すくなると考えられる。MFCC の計算の概要を以下に示す。音声波形 $y(t)$ をフレーム長 0.025[s]、シフト長 0.01[s] で切り出し、ハン窓を掛けた。

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 $S(\omega)$ を求める。

$$S(\omega) = |Y(\omega)| \quad (7)$$

$S(\omega)$ 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 を用いることとしたため、20 個の三角窓を合わせたフィルタを使用した。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[7]。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。

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 の特徴量の評価を行う。音声データは 1 年目の数納氏による卒業研究「睡眠時の呼吸音の収録及び音環境調査」[8]で収録したデータを使用している。

機材は図 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 & Kjaer TYPE4231)を使用した。



図 1 使用機材

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。時間毎のピーク周波数の変化の波形を図 2 に示す。

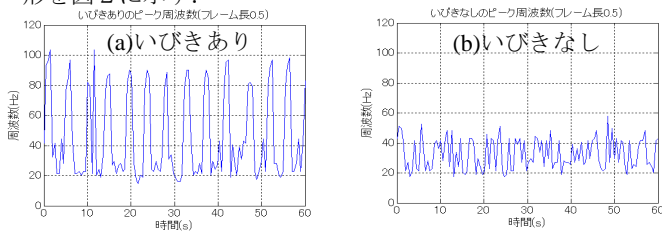


図 2 時間毎のピーク周波数の変化

今回抽出されたピーク周波数は大体 70Hz~120Hz の範囲にいびき音が出ており、70Hz 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化が大きく、ノイズの部分は時間毎の変動が。このことから、ピーク周波数からいびきの判別は可能であると考えられる。次にエンベロープ波形を図 3 に示す。

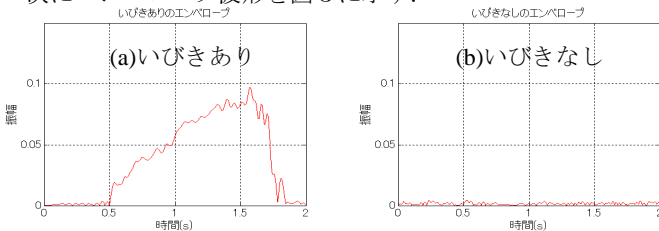


図 3 エンベロープ波形

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、エンベロープ波形からいびきの判別は可能であると考えられる。最後に MFCC のグラフを図 4 に示す。

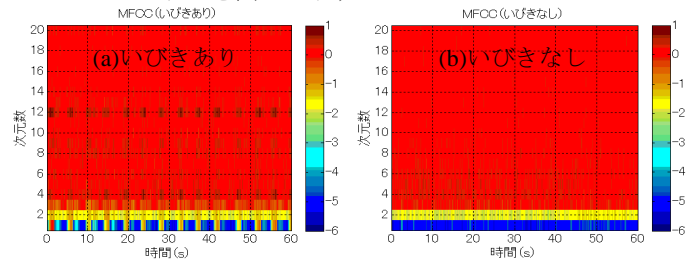


図 4 MFCC

図 4 は、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である。高次成分の違いは分からないが、低次成分は若干ではあるが違いが出ていた。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った。これは MFCC の性質として有音声だと 1 次元目の値が大きくなり、無音声だと値が小さくなるためである[9]。目視による確認だけでは評価が難しいため、詳しい結果は共同研究者の本学・計算機構成研究室所属の加科氏による卒業研究「呼吸音検出アルゴリズムの構築～判別器の作成及び評価～」で報告する。

5. まとめと今後の課題

ピーク周波数を出すときのフレーム長は、短い間隔よりも少し長めにした方が周波数分解能が上がりいびき以外に起因する成分を分離して、判別しやすくなるのが分かった。しかし最適なフレーム長は環境やいびきの種類によって変わる可能性があり、今後も調査が必要である。エンベロープでは振幅が大きくなるにつれて振幅の変化も大きくなった。いびき音が小さくても同様の傾向が表れるのか検証が必要である。MFCC は機械学習による判別器に読み込ませたところ、分類率は 0.85 であった。誤分類されたデータがあった原因としては雑音といびきの波形の形が似ていた、いびき音が小さくて認識されていない等が考えられる。また、短いフレーム長で切り出していった波形から求めるため、フレーム長によって計算時間や結果が変化する。そのため最適なフレーム長を検討する必要がある。

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色々な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. 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] 飯塚祐太「睡眠時の呼吸音検出に関する研究」工学院大学卒業論文，2011
- [6] 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
- [7] 古井貞照著「新音響・音声学」近代科学社，2006
- [8] 数納聖洵「睡眠時無呼吸症候群の簡易診断システムの構築～睡眠時における音環境調査～」工学院大学卒業論文，2012
- [9] 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