［研究論文］

画像と音声情報を用いたうつ病 CAD の高性能化

和田昇太 1・牧優太 1・安倍和弥 2・武尾英哉 2・永井優一 3

1 博士後期課程電気電子工学専攻
2 電気電子情報工学科
3 国立がん研究センター東病院

Developing High Performance CAD for Depression by Employing Image and Voice Information

Shota WADA[1], Yuta MAKI[1], Kazuya ABE[1], Hideya TAKEO[1], Yuichi NAGAI[2]

Abstract

The number of depressed people is on the rise in Japan. In 2011, depression was added to the four major diseases, making it one of the five major diseases. This suggests the need for reliable detection of depressed people. Currently, the diagnosis of depression is mainly based on the interview. However, this lacks objective indicators, and some have questioned the accuracy of the diagnostic results. In order to solve this problem, we are developing an AI-based system that calculates the degree of certainty that a patient is depressed by focusing on facial images and voice information. The purpose of this study is to use this as an objective indicator to assist in the diagnosis of depression. This paper describes the construction of a voice discriminator that constitutes the system. A total of 42 features were created based on the patient's speech and the characteristics of the depressed person. The SVM was used to create and evaluate the discriminator, and the accuracy was 90% for the female data. However, the accuracy for male data was about 55%. Hence, these features were selected by 18 feature selection methods. The discriminator was made and evaluated again, and the accuracy was 80%, even for male data.

Keywords: Depression, Voice information, Machine learning, SVM, Feature selection

1. はじめに

我が国において，うつ病患者は増加傾向にある[1)]．また，平成19年から21年の警察庁統計における自殺の動機として各年うつ病が原因であるものが4割以上を占め，厚生労働省では自殺対策として，うつ病対策を重点的に行っている[2)]．こうした背景もあり，2011年には，従来の「がん」，「脳卒中」，「急性心筋梗塞」，「糖尿病」からなる4大疾病に，新たにうつ病を含む「精神疾患」が追加され5大疾病となり，国民の健康保持のため広範かつ継続的な医療の提供が必要な疾病であるとされた[3)]．

現在のうつ病診断は，DSM-5（精神障害の診断と統計マニュアル第 5 版）や ICD-10（国際疾病分類第 10 版）などの診断基準に照らし合わせ行われるものである．しかしながらこれが客観性を欠き，うつ病者の発見を阻害しているという指摘もあり，うつ病判別に有効なバイオマーカーについての研究が行われている[4)]．

例えば人間の発話音声を対象として分析し，バイオマーカーとして用いた研究として，和家ら[5)]は，収録環境を統一したうえで様々な課題に沿って収録した音声を分析し，従来から検討されてきた音声のエネルギーなどの有効性を確認し，また時間内に想起された単語数といった新たなバイオマーカーを抽出して有効性を示した．しかしながらこれらの手法には，文章読み上げなど患者側の負担が考えられる．

我々もうつ病判別システムの開発を行ってきた[6)7)8)]．これは AI 技術と画像工学技術を組み合わせうつ病の確信度を算出することで，医師に客観的な指標を提示し，診断の

参考にしてもらうことを目的としたものである. 本研究においては, 新たに検討した, 音声を用いたうつ病判別器について述べる, 最終的に前述の画像工学技術を用いた判別器と統合し, システム全体としての判別精度向上を実現する. 発話音声については定型文読み上げなどを用いない自由発話の音声を用いて判別器の検討を行う. この 2 点が新規な点である.

本論文の構成としては, 2 章で研究に用いたデータについてとその分析結果について整理し述べる. 第 3 章では判別器の作成・評価方法とうつ病者に見られる特徴をグループ分けしそれぞれについて検討した特徴量の算出方法について述べる. 4 章では, 3 章の内容についての結果を示し, 5 章ではその考察を述べる. 6 章で現在の取り組みについて, 7 章で総まとめとする.

2. データ

2.1. データの準備

本研究のデータは, 動画投稿サイトよりうつ病者・健常者のそれぞれについて, 男女各 10 名ずつの計 40 名の動画を収集した(表 1). 実際の問診を想定し動画の内容は, 自己紹介など自分について自由に話しているものとした. 動画からサンプリング周波数 44,100[Hz], 16[bit]で音声を抽出し, モノラルオーディオに変換したものから 240[s]に相当する 10,584,000[sample]を切り出し利用した. またうつ病者・健常者のラベル付けは, 本来ならば医師の協力を得て行うことが望ましいが現状不可能であるため, 自身が動画内などでうつ病を罹患していると申告している者をうつ病者, そうでない者を健常者とした.

表 1. 収集し研究に用いたデータの総数

	うつ病者	健常者	合計
男性	10	10	20
女性	10	10	20
合計	20	20	40

2.2. データの分析

前節で準備したデータを視覚・聴覚的に評価した. うつ病者と健常者では,

・発話量に乏しい
・発声量に乏しい
・声の抑揚に乏しい

などの特徴がみられた.

これらの観察結果と先行研究などの文献 [5)9)] とを照らし合わせたうえで, うつ病者の発話音声に特有の特徴をグループ分けし以下表 2 にまとめた.

一例としてグループ 2 について, うつ病者と健常者の音声波形を以下図 1, 図 2 に示す. うつ病者は振幅が小さく声量が小さい傾向がみられた.

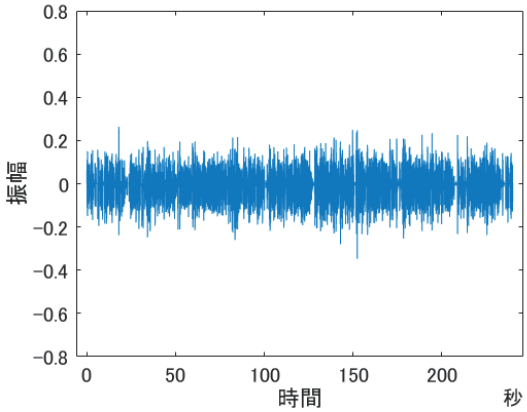

図 1. うつ病者の音声波形の例

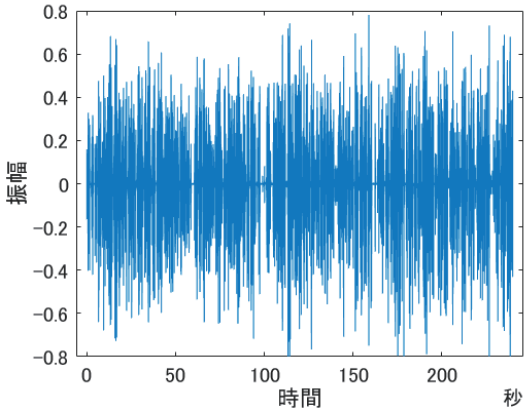

図 2. 健常者の音声波形の例

表 2. うつ病患者の特徴

グループ	発話音声の特徴	うつ病患者の場合
1	発話速度	1. 口数が少ない
		2. 話し方が単調で途切れ途切れ
		3. 動作が遅い
		4. 思考速度が遅い
2	発声量	1. 声量が小さい
3	抑揚	1. 抑揚が乏しい
		2. 活気がない
		3 感情が動かない

3. 手法

判別器は 2.1 節で述べた男女計 40 名のデータを用いて SVM(Support Vector Machine : サポートベクターマシン)及び LOOCV(Leave-One-Out Cross-Validation : 一個抜き交差検証)で作成する. それにあたり, 表 2 にまとめたうつ病者の特徴をもとにグループごとに複数の特徴量を作成した. 学習済みの判別器は, 0〜1 の範囲でうつ病の重症度を出力する. 判別器の評価のため, 重症度の閾値を 0.5 とし, それ以上をうつ病, 未満を健常者と判別するようにしたうえ判別精度, 式(1)のようにして評価を行った.

判別器は男女の性差を考慮するため, 及び判別精度向上のための特徴量選択による次元圧縮をより有効的に行うために男女でデータを分けて, 男性用判別器, 女性用判別器の 2 つを作成・評価した.

$$判別精度 = \frac{真陽性+真陰性}{全体} \quad (1)$$

3.1. 特徴量の作成（グループ 1 の定量化）

発話速度に関連した特徴を表 2 中グループ 1 とした．うつ病患者には表 2 中の 1-1. 口数が少ないという傾向がある．また表 2 中 1-2. 話し方が単調，1-3. 動作が遅い，1-4. 思考速度が遅いなどの特徴と相まってこの傾向を強めていることが考えられる．そこで，聴覚評価により，音声 240[s]に相当する，10,584,000[Sample]の発話・非発話区間をラベル付けし，全サンプルにおける発話サンプルの割合を特徴量とし定量化した．以下図 3，図 4 の有色部分は，それぞれ音声 60[s]に相当する 2,646,000[Sample]中の有声部分である．一発話の継続時間や 1 秒以上のつまり（ポーズ）回数も計算し特徴量として用いた．

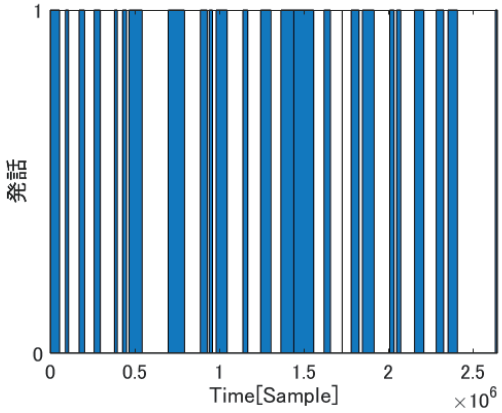

図 3. うつ病者の発話量

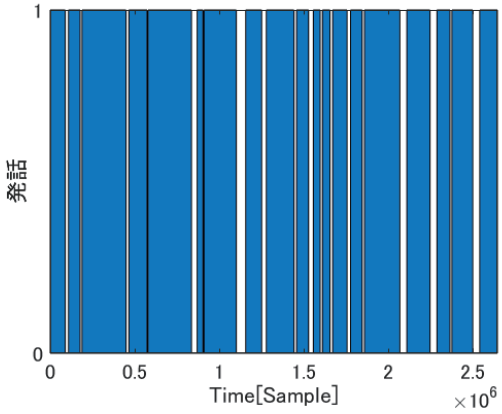

図 4. 健常者の発話量

3.2. 特徴量の作成（グループ 2 の特徴定量化）

発声量について，健常者とうつ病者で差が見られた．具体的には，健常者の方が発声量が大きく，逆にうつ病者においては小さいといった具合である（図 1，図 2）．発声量を以下(2)式，(3)式の様に計算し特徴量として定量化した．音声波形をフーリエ変換したときのパワースペクトルの例を以下図 5，図 6 に示す．音声波形の変動成分はうつ病者において小さい傾向にあった．

$$振幅・FFT：RMS[Pa] = \sqrt{\frac{1}{t_2-t_1}\int_{t_1}^{t_2} x^2(t)dt} \quad (2)$$

$$信号パワー：p[w/m^2] = \frac{1}{\rho cT}\int x^2(t)dt \quad (3)$$

ρ：空気の密度：$1.14[kg/m^3]$
c：空気中の音速：353[m/s]
T：音源の長さ：[s]
$x^2(t)$：音圧[Pa]

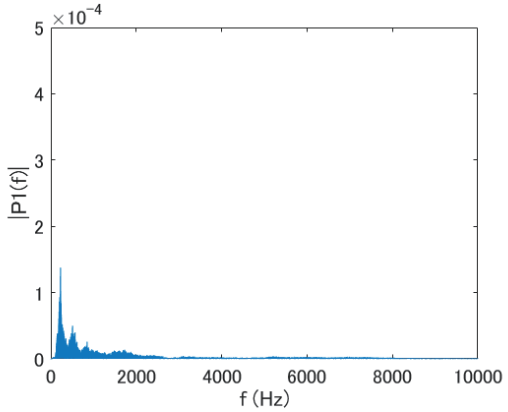

図 5. うつ病者の音声パワースペクトル例

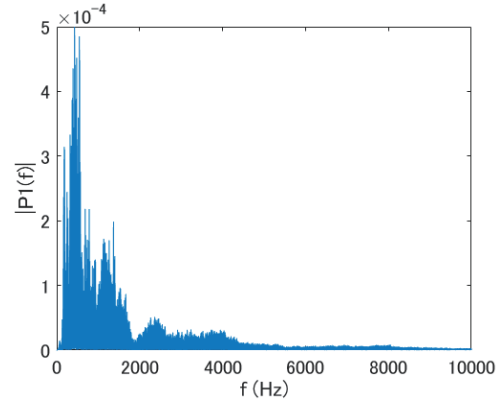

図 6. 健常者の音声パワースペクトル例

3.3. 特徴量の作成（グループ 3 の特徴定量化）

うつ病者における発話音声の特徴として，抑揚の消失があげられる．表 2 中の 3-1. 抑揚に乏しい，3-2. 活気がないといった傾向をグループ 3 に分類し定量化した．特徴量には音声信号から wavesurfer[10]を用い抽出した基本周波数 F0 を用いた．同ツールでは，RAPT（A Robust Algorithm for Pitch Tracking）というアルゴリズムを用いることで，精度の高い F0 抽出を可能にしている．0.01[s]毎に抽出した F0 と，その変化量ΔF0 の時系列波形から，最大・最小値や平均値，分散などの基本統計量を計算し，特徴量として用いた．以下図 7 に抽出した F0 の時系列波形を示す．

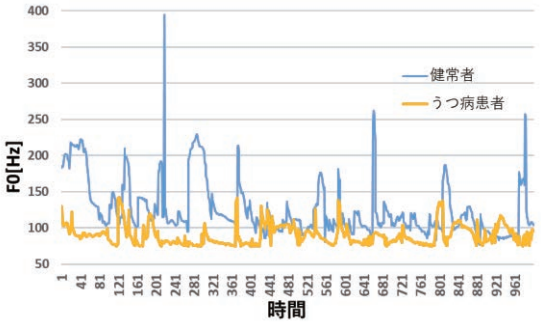

図 7. 抽出した F0 の時系列波形

3.4. 特徴量選択

判別精度向上の為に本節迄の手法により作成した特徴量 42 個を選択・圧縮し，判別器を作成後評価した．以下にその手法を述べる．

3.4.1. 選択手法①

入力された特徴量を表 3，表 4 に示す計 18 の手法によ

り選択する．この過程は 2 段階に分けて行う．まず 1 段目では，表 3 に示す 15 の特徴量選択方法(アルゴリズム)を用いて選択を行う．これらの複数のアルゴリズムにより 7 回以上選ばれた特徴量 X を次段に入力する．2 段目では入力された特徴量 X を表 4 の 3 つのアルゴリズムを用いて再度選択し 3 回以上選ばれた特徴量を採用する．

これら一連の流れを手法①として，概略図を以下図 7 に示す．表 3，表 4 中の各方法の詳細な説明については，次の資料を参照されたい [11]．

3.4.2. 選択手法②

入力された複数の特徴量の組合せ全通りを入力として SVM による判別器作成と LOOCV による評価を行い，判別精度が最大になったときの特徴量を選択する．これを手法②とし，概略図を以下図 8 に示す．

3.4.3. 各手法の組合せ

本項までに述べた手法①，手法②，及びランダムフォレストを組み合わせて特徴量選択を行い，最終的な判別器の精度を評価した．大幅に特徴量の次元圧縮ができる手法①を 1 段目に用いることを念頭に，以下の組合せについて検討を行った．1)，2)では，それぞれの選択手法を単体で用いた場合について．3)の組合せの 1 段目を手法①に置き換えたのが 4)である．

1)ランダムフォレストのみ
2)手法①のみ
3)ランダムフォレスト ＋ 手法②
4)手法① ＋ 手法②

表 3. 1 段目の特徴量選択方法

段階	方法	
1	Filter	分散分析
		χ2乗統計量
		最小所調整特徴選択
		F検定
		ラプラシアンスコア
		相関
		近傍成分分析
		RelieF
	Wrapper	
	Embedded	線形判別分析分類器
		ECOC(誤り訂正出力符合)
		線形会期モデル
		リッジ回帰モデル
		ElasticNet正則化
		Lasso正則化

表 4. 2 段目の特徴量選択方法

段階	方法	
2	Decision tree	分類木
		分類木のアンサンブル
		ランダムフォレスト

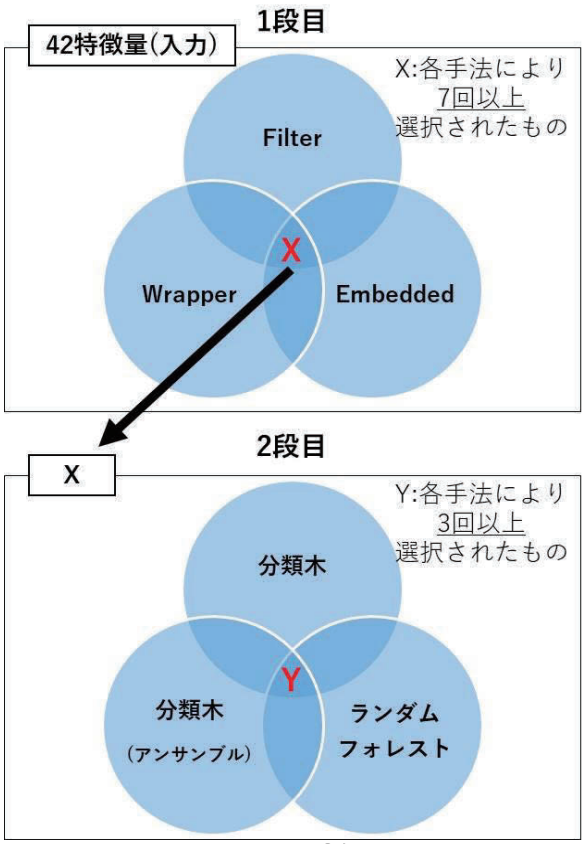

図 7. 手法①概略

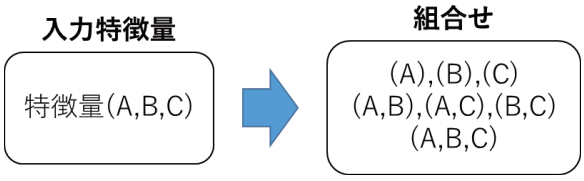

図 8. 手法②概略

4. 結果

3.1～3.3 節で述べた手法で算出した特徴量の一部分布を以下 4.1～4.3 節に示す．4.4～4.5 節では特徴量選択の結果について述べる．

4.1. グループ 1 の特徴量について

音声 240[s]相当の 10584000[Sample]における発話サンプルの割合を発話量と定義した．一例として発話量の最大値が 1，最小値が 0 となるように正規化したものをヒストグラムにし以下図 9 に示す．また発話中に 1 秒以上詰まる回数について，同様に正規化後ヒストグラムにしたものを以下図 10 に示す．

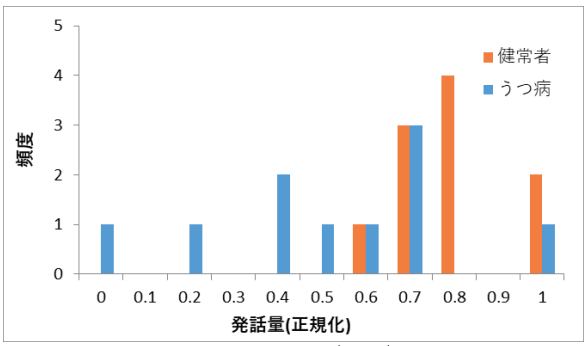

図 9. 発話量(男性)

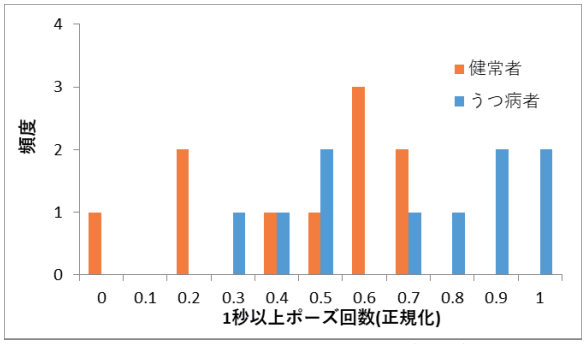

図 10. 1 秒以上ポーズ回数(男性)

4.2. グループ 2 特徴量について

3.2 において定義した式(2)より信号のパワーを計算した．これは表 2 中特徴グループ 2 の声量を定量化したものに相当する．一例として以下図 11 に示す．

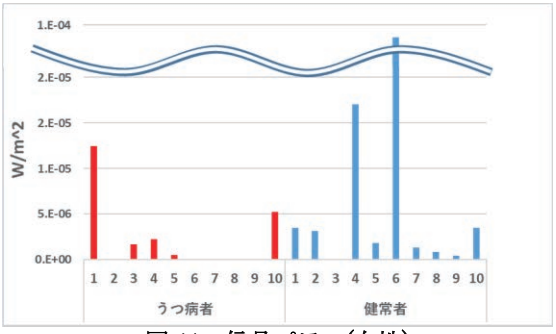

図 11. 信号パワー(女性)

4.3. グループ 3 の特徴量について

ΔF0 の最大値，及び分散について計算し，それらの最小値が 0，最大値が 1 となるように正規化したうえでヒストグラムにまとめたものを一例として以下図 12，図 13 に示す．これらは，特徴量グループ 3 の 1. 抑揚の大きさに相当する．

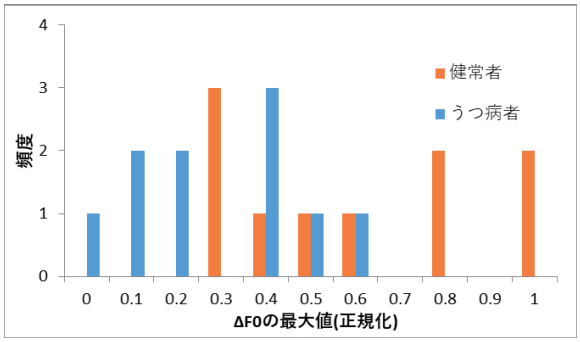

図 12. ΔF0 の最大値(女性)

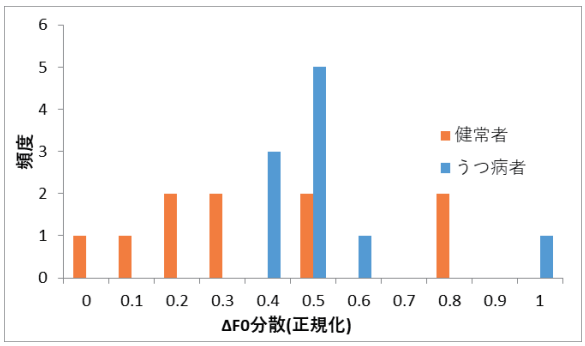

図 13. ΔF0 の分散(女性)

特徴グループ 1，2，3 の計 42 個の全ての特徴量を用いて，SVM で学習したときの判別精度は，男性判別器で 55%，女性判別器で 90%であった．

4.4. 特徴量選択

3.4.3 項で述べた各手法の組合せ(1)～(4)により選択された特徴量の個数と，それを用いて SVM，LOOCV で学習・評価を行ったときの判別精度を以下表 5 に示す．このとき各判別器には LOOCV により 20 のデータ(訓練データ 19，テストデータ 1)が学習に用いられる．

表 5. 選択された特徴量の個数と判別精度

各手法の組合せ	データ	選択された特徴量	判別精度
(1) ランダムフォレストのみ	男性	8個	70%
	女性	8個	90%
(2) 手法①のみ	男性	6個	75%
	女性	7個	75%
(3) ランダムフォレスト ＋ 手法②	男性	3個	80%
	女性	3個	90%
(4) 手法① ＋ 手法②	男性	3個	90%
	女性	2個	90%

4.5. 特徴量選択の有用性

特徴量 42 個を全て用いたときの判別器の判別精度と，手法の組合せ(4)により選択された特徴量を用いたときの判別精度を以下表 6 に示す．男性判別器について，35 ポイントの大幅な精度向上が見られた．

表 6. 本手法で得られた判別精度まとめ

	全42特徴量を使用	特徴量選択法(4)
男性	55%	90%
女性	90%	90%

5. 考察

表 6 より，特徴量選択法によって男性判別器，女性判別器でともに 90%の判別精度を得ることができた．表 5 より今回の特徴量選択手法では，最終的に 3 つ程度の特徴量が選択され判別器を作成している．よって判別に寄与しない特徴量が十分に除外され精度が最大化できたと考えられる．その一方で，未知データに対する汎化性能が低下する可能性があり，新たにデータを加えた場合の判別精度についても検証する必要があると考えられる．汎化性能低下がみられた場合には．例えば表 2 に示した特徴グループに対応した特徴量を，各グループから必ず複数個以上の特徴量が選ばれるようにするなどの方法で汎化性能を担保するなどの検討が必要であると考えられる．

また全特徴量 42 個を用いたときの男性判別器の判別精

度が著しく低いことからも，よりうつ病者の特徴をとらえた特徴量作成について引き続き検討することが最優先事項であると考えられる．

6. 今後の展望

今後の展望として，現在検討していることについて述べる．

6.1. 特徴グループ 4 の設定

音声の評価を改めて行ったところ，うつ病者は健常者に比べて声がかすれているように聞こえる傾向が見られた．そのため，これらの特徴を，以下表 7 に示す特徴グループ 4 に設定した．音声合成システムにおいて，かすれ声を合成する際に非周期的な信号を加えることから，発話音声中に含まれる非周期成分の割合を，声分析変換合成システム”WORLD”[12]を用いて 5[ms]毎に計算した．以下図 14，図 15 にその時系列波形を示す．またその平均値を計算することで特徴量としてかすれを定量化した．健常者とうつ病者で，平均値の分布をヒストグラムにしたものを以下図 16 に示す．今後はこれらの特徴量も詳細に検討し，判別器作成に用いる予定である．

表 7. 特徴量グループ 4 の設定

グループ	発話音声の特徴	うつ病患者の場合
4	その他の特徴	1.悲観的
		2.劣等感や無力感
		3.声がかすれる
		4. 周囲や他者への興味関心の喪失

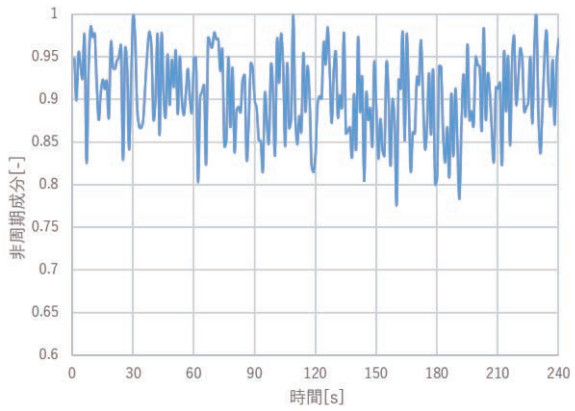

図 14. 非周期成分(うつ病者)

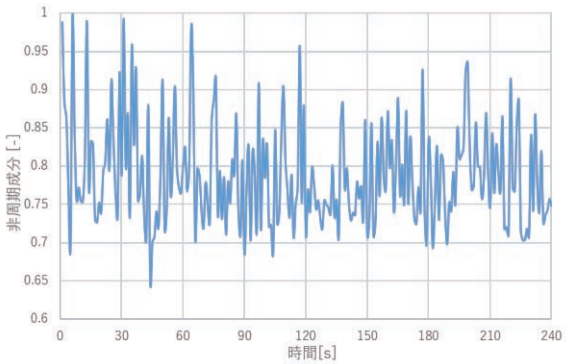

図 15. 非周期成分(健常者)

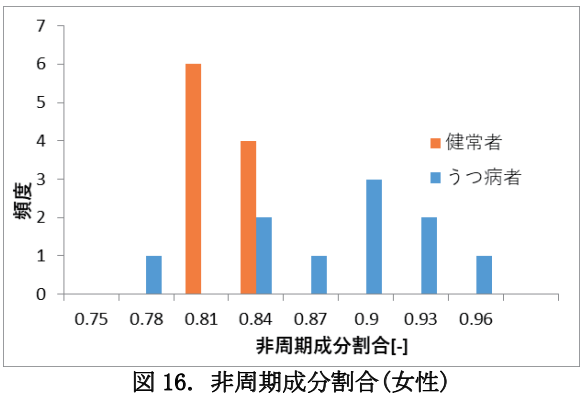

図 16. 非周期成分割合(女性)

6.2. 発話内容に着目しうつ病判別器を作成

うつ病者の特徴として表 7 中 4-1.悲観的になる，4-4.他人への関心の低下があげられる．そのため，この特徴が発話内容に影響を及ぼす可能性を言語分析的観点から分析し，判別器を作成することを検討している．具体的には，発話音声の内容を全て文字起こししたうえで，例えばポジティブな単語には加点，逆にネガティブな単語には減点といったスコア付けを行い何か傾向が見られないか分析する．現状は発話内容の文字起こしを行っており，以下図 17 は女性のうつ病．健常者 8 人ずつの発話を全てひらがなで文字起こしたときの文字数を示している．ここから前述の言語分析や，モーラ拍を定義し，グループ 1 の特徴量として加えるなどを検討する．図 18 は発話中に発現する形容詞に対してスコア付けを行った例である．

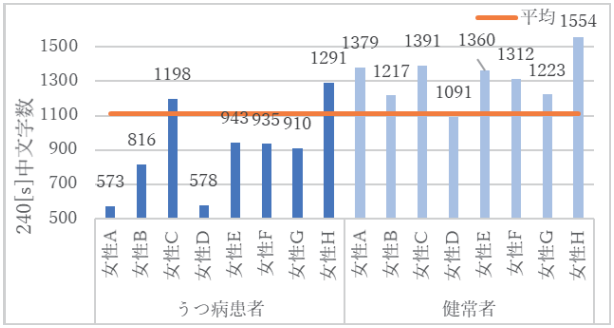

図 17. 女性音声の発話量

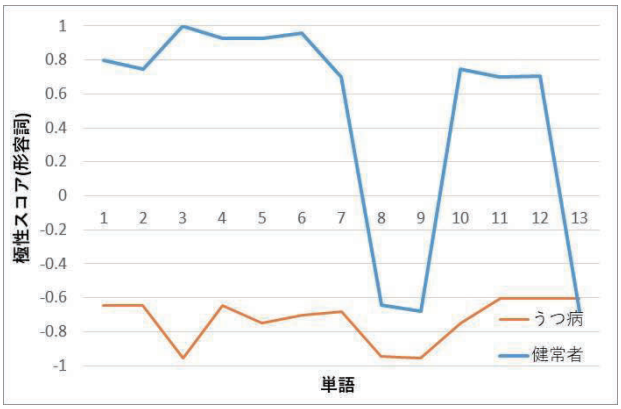

図 18. 形容詞に対するスコア付けの例

7. まとめ

我々が提案するうつ病判別システムに組み込むべく，うつ病の音声判別器の作成と高性能化について検討した．特徴量選択の後，作成・評価を行った判別器の判別精度は，男性データ，女性データともに 90%となった．今後は 6.1

節で述べたような，より判別に有意な特徴量の追加による判別精度の向上や，6.2 節で述べた新たな判別器の作成，特徴量選択の最適化による判別器の汎化性能の向上についての検討が必要であると考えられる．

参考文献

[1] 厚生労働省：うつ病対策について，
https://www.mhlw.go.jp/kokoro/nation/dyp.html
(2020 年 7 月 31 日閲覧)
[2] 厚生労働省：自殺・うつ病等対策プロジェクトチームとりまとめについて，
https://www.mhlw.go.jp/seisaku/2010/07/03.html
(2020 年 7 月 31 日閲覧)
[3] 厚生労働省：5 疾病・5 事業について，
https://www.mhlw.go.jp/file/05-Shingikai-10801000-Iseikyoku-Soumuka/0000139231.pdf
(2020 年 7 月 31 日閲覧)
[4] 厚生労働省：研究内容と成果(うつ病に関する例)，
https://www.mhlw.go.jp/shingi/2009/08/dl/s0806-16b_0013.pdf (2020 年 7 月 31 日閲覧)
[5] 和家　尚希，鈴木　雅之，長野　徹　他：精神疾患診断補助に有効な発話課題と音声特徴に関する検討．信学技報　SP2014-133(2015)
[6] 牧優太 和田昇太 安倍和弥 他： 画像認識技術によるうつ病診断の定量化 第 38 回 日本医用画像工学会大会 (JAMIT2019) 予稿集 ， OP1-17 ， 2019
[7] Maki Y, Wada S, Abe K, et al.: Developing High Performance CAD for Depression by Integrating Multiple Classifier Systems. Computer Assisted Radiology and Surgery 34rd International Congress and Exhibition, CARS2020,S204-S205, 2020
[8] Wada S, Maki Y, Abe K, et al.: Developing High Performance CAD for Depression by Employing Image and Voice Information. Computer Assisted Radiology and Surgery 34rd International Congress and Exhibition, CARS2020, S203-S204, 2020
[9] 忠井俊明：ようこそ精神医学へ．ミネルヴァ書房，2003
[10] http://www.speech.kth.se/wavesurfer/
[11] MathWorks Introduction to Feature Selection
https://jp.mathworks.com/help/stats/feature-selection.html?lang=en (2020 年 7 月 31 日閲覧)
[12] http://www.kki.yamanashi.ac.jp/~mmorise/world/