

筑波大学

博士（医学）学位論文

精神疾患における  
音声バイオマーカーの検討

2017

筑波大学大学院博士課程人間総合科学研究科

田口 高也

## 目次

略語 .....	6
研究背景 .....	7
1. うつ病の疫学 .....	8
2. うつ病の診断指標 .....	8
3. 音声と音響特徴量について .....	10
4. MFCC .....	10
5. 音声と抑うつ状態に関する報告 .....	11
6. 本研究の目的 .....	12
研究 1 音響特徴量によるうつ病患者と健常者の判別 .....	13
1. 目的 .....	14
2. 対象と方法 .....	14
2.1. 対象 .....	14
2.2 音声の録音 .....	15
2.3. 音響特徴量の抽出 .....	15
2.4. 統計解析 .....	17
3. 結果 .....	18
3.1. 各群の音響特徴量の差の検討 .....	18
3.2. MFCC による判別分析 .....	18
3.3. 各群における周波数スペクトルの差の検討 .....	18
3.4. 音響特徴量と QIDS-SRJ の相関解析 .....	19
3.5. MFCC 2 における重回帰分析 .....	19
3.6. 言語流暢性課題前後の音声における反復測定の一元配置分散分析 .....	19
4. 考察 .....	20
4.1. MFCC 2 の変化はなぜ起こるのか .....	20
4.2. 基本周波数 .....	20
4.3. 小括 .....	21
研究 2 治療による抑うつ症状と音響特徴量の変化 .....	22
1. 目的 .....	23
2. 対象と方法 .....	23
2.1. 対象 .....	23

2.2. 抑うつ症状の評価 .....	23
2.3. 音声の録音・抽出 .....	23
2.5. 統計解析 .....	24
3. 結果 .....	24
3.1. ECT 前後における QIDS-J スコアの対応のあるサンプルの t 検定と相関解析 .....	24
3.2. ECT 前後における音響特徴量の相関解析と対応のあるサンプルの t 検定 .....	24
3.3. ECT 前後における QIDS-J スコアの変化と音響特徴量の変化の相関解析 .....	24
4. 考察 .....	25
4.1. ECT 前後の抑うつ症状の変化 .....	25
4.2. ECT 前後の音響特徴量の変化 .....	25
4.3. ECT 前後での QIDS-J スコアの変化と音響特徴量の変化の関係 .....	26
4.4. 小括 .....	26
研究 3 うつ病患者、統合失調症患者、健常者の音声の比較 .....	27
1. 目的 .....	28
2. 対象と方法 .....	28
2.1. 対象 .....	28
2.2 音声の録音・抽出 .....	29
2.4. 統計解析 .....	29
3. 結果 .....	29
3.1. 各群の音響特徴量の比較 .....	29
3.2. 音響特徴量と QIDS-SRJ、BPRS との相関 .....	29
3.3 統合失調症群のうち抑うつが中等度以上ある群とない群の差の比較 .....	30
4. 考察 .....	30
4.1. 統合失調症の抑うつ症状 .....	30
4.2. 小括 .....	31
総合考察 .....	32
1. 研究成果のまとめ .....	33
2. 研究成果の解釈 .....	33
3. 音声はバイオマーカーになりえるのか .....	34
4. 研究の限界と今後の方向性 .....	35
結論 .....	38

謝辞 .....	39
文献 .....	40
図表 .....	45
図 1. 音声の成り立ちと音響特徴量の抽出 .....	46
図 2. 使用したタブレット端末と受話器型ヘッドセット .....	47
図 3. 使用したマイクロフォンの周波数特性 .....	48
図 4. うつ病群とコントロール群における MFCC 2 .....	49
図 5. MFCC 2 の RECEIVER OPERATING CHARACTERISTIC (ROC) 曲線 .....	50
図 6. うつ病群とコントロール群の音声スペクトルの比較 .....	51
図 7. MFCC 2 のうつ病群とコントロール群における差分、および MFCC 0 から 3 の離散コサイン変換における基底 .....	52
図 8. 語流暢性課題の前後における MFCC 2 .....	53
図 9. コントロール群、統合失調症群、うつ病群の比較 .....	54
表 1. 音響特徴量の例 .....	55
表 2. 対象 .....	56
表 3. うつ病群とコントロール群の比較 .....	57
表 4. MFCC 2 における判別分析の結果 .....	58
表 5. MFCC 2 の重回帰分析 .....	59
表 6. 対象 .....	60
表 7. ECT 前後の QIDS-J .....	61
表 8. 対象 .....	62
参考文献 .....	63

## 略語

AUC	Area under curve	曲線下面積
BDNF	Brain-derived neurotrophic factor	脳由来神経栄養因子
CO <sub>2</sub>	Carbon dioxide	二酸化炭素
COPD	Chronic Obstructive Pulmonary Disease	慢性閉塞性肺疾患
CRP	C-reactive protein	C 反応性蛋白
DSM	Diagnostic and Statistical Manual of Mental Disorders	精神疾患の診断・統計マニュアル
DALY	Disability-Adjusted Life Year	障害調整生命年
ECT	Electroconvulsive therapy	電気けいれん療法
FFT	Fast Fourier Transform	高速フーリエ変換
ISO	International Organization for Standardization	国際標準化機構
LPC	Linear Predictive Coding	線形予測符号
MFCC	Mel Frequency Cepstral Coefficients	メル周波数ケプストラム係数
PCM	Pulse Code Modulation	パルス符号変調
PLP	Perceptual Linear Prediction	知覚的線形予測
QIDS-J	Quick Inventory of Depressive Symptomatology, Japanese version	簡易抑うつ症状尺度 日本語版
QIDS-SRJ	Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version	自己記入式簡易抑うつ症状尺度 日本語版
ROC	Receiver Operating Characteristic	受信者動作特性
WHO	World Health Organization	世界保健機関
YLD	Years Lived with Disability	障害生存年数
YLL	Years of Life Lost	損失生存年数

## 研究背景

## 1. うつ病の疫学

うつ病は、精神科領域における重要な疾患であるが、患者数の増加や社会的影響から、もはや精神科領域だけでなく、社会全体の問題として認識されている。世界保健機関（WHO）が2017年に発表(WHO, 2017a)したように、2015年時点でうつ病患者は世界で3億人を超えたと推定され、2005年から18%以上増加している。また、WHOはthe Global Health Estimates 2015(WHO, 2017b)

において、疾病により障害を余儀なくされた時間を表す障害生存年数（Years Lived with Disability: YLD）の大きさを順位付けしている。ここで、うつ病の順位は2000年の報告で鉄欠乏性貧血に次ぐ第2位であったものが、2015年には第1位となっている。さらに、疾病のために予測されるよりも早く死亡することの損失年数を表す損失生存年数(Years of Life Lost: YLL)とYLDとの和である障害調整生命年（Disability-Adjusted Life Year: DALY）が疾病損失を総合的に評価する指標として用いられているが、DALYにおいても2000年にうつ病は17位、2015年には13位と主要な疾患に含まれ、かつ順位を上げており、その社会的負担が重要視されるようになってきている。

2006年の川上らの報告によると、我が国のうつ病の12カ月有病率は1~2%、生涯有病率は3~7%であり、欧米の12カ月有病率1~8%、生涯有病率3~16%に比べると低い。また一般的に女性、若年者に多いとされるが、日本では中高年でも頻度が高いとされる(川上, 2006)。しかし、先の2015年のWHO報告では、日本のうつ病の推計患者数は506万人と厚生労働省による平成26年の患者調査(Ministry of Health Labour and Welfare (Japan), 2014)における気分障害通院患者統計112万人の5倍にあたり、実際は我が国のうつ病潜在人口はかなり多く、かつ増加していることが推測される。以上から、我が国においても、また世界的にもうつ病の早期診断と有効な治療法の開発は喫緊の課題である。

## 2. うつ病の診断指標

うつ病の診断は他の精神疾患と同様主に疾患を指定した臨床症状の厳密な記述と吟味によってなされるが、現在は「精神疾患の診断・統計マニュアル（Diagnostic and Statistical Manual of Mental Disorders; DSM）」の登場により診断基準を用いた操作的診断の存在が大きくなった(American Psychiatric Association. and American Psychiatric Association. DSM-5 Task Force., 2013)。DSM-5におけるうつ病（Major Depressive Disorder: うつ病性障害）は、「抑うつエピソードが同じ2週間に存在し、病前の機能からの変化を起こしており、他の原因で説明がつかず、臨床的に意味のある苦痛、または社会的、職業的、または他の重要な領域における機能の障



害を引き起こしている」と定義される。抑うつエピソードを満たすためには、①抑うつ気分、②ほとんどすべての活動における興味または喜びの喪失、のいずれかが少なくとも2週間存在し、この他に、③食欲または体重、④睡眠、⑤精神運動性活動の変化、⑥気力の減退、⑦無価値観または罪責感、⑧思考、集中、決断の困難、⑨繰り返される死についての考え、または自殺念慮・計画・企図、のうち少なくとも4つを体験していなければならない。

うつ病の診断補助として、診断基準を元にした心理尺度も作成されている。例えば簡易抑うつ症状尺度 (Quick Inventory of Depressive Symptomatology; QIDS) は DSM-IV の抑うつエピソードを元に作成されており、そのスコアで重症度の評価を行うことができる (Rush et al., 2003)。しかし、このような心理尺度は評価者の主観に大きく影響を受けるため、より客観的な評価方法の開発も望まれている。

近年、血液中の物質や遺伝子など、うつ病の生物学的要因と関連する指標 (バイオマーカー) を見出そうとする研究が盛んとなってきている (Strawbridge et al., 2017)。しかし、上述のようにうつ病の疾患概念は症候学的に規定されており、病因は多様で複雑な症候群であるという考え方が一般的となっているため (Strawbridge et al., 2017)、疾患を確実に検出する検査の開発は困難と考えられる。これまで、コルチゾールなどの内分泌ホルモン (Fischer et al., 2017)、セロトニンやノルアドレナリン、ドーパミンなどの神経伝達物質 (Coplan et al., 2014; Kaufman et al., 2016)、脳由来神経栄養因子 (BDNF) (Molendijk et al., 2014)、高感度 CRP (C 反応性蛋白) や炎症性サイトカイン (Haapakoski et al., 2015)、さらに各種遺伝子多型 (Dunn et al., 2015)、構造・機能画像 (Fu et al., 2013; Wise et al., 2014) に至るまで、様々なものが指摘されてきた。しかし、いずれもバイオマーカーとして臨床での有用には至っておらず、実際の症状との関連も明確な報告は少ない。

一方、より簡便な診断補助指標として、患者の非言語的な情報も臨床現場では重要な症候とされてきた。一般に精神医学的な症候は、患者の話す情報の他にも、態度や姿勢、身体的状態などの非言語的情報や感情も診断の参考とされる (Ladd, 1980)。この非言語的情報のうち音声の果たす役割は大きい。DSM-5 のうつ病の症候解説においても、重要症候の一つである精神運動制止の例として、「声量、抑揚、会話量、または内容の豊かさの減少」が挙げられている。これらは「他人にもわかるほど十分に重症なものではなければならず」と注釈されたうえで、医師が経験的に判断することが想定されている。実際にもうつ病患者の声がこもって聞こえ、内容が聞き取りづらいということがよく経験される。では、音声の特徴は、客観的にはどのように理解され、うつ病とどのように関連するだろうか。

### 3. 音声と音響特徴量について

音声は、声帯が振動することで発生するが、それが喉頭、咽頭、鼻腔、副鼻腔、口腔で共鳴し、舌の動きや口の開き方などにより「音声」として形成される。音声には言語的な情報のみならず、多くの情報が含まれる。例えば、声の音量や震えの有無、話速、発声しない時間などによって、電話による会話の場合であっても、話者の感情を推し量ることも可能である。実際にはそれだけではなく、声の雰囲気も我々は無意識に受け取っていることが多い。これは我々が言語コミュニケーションの経験の中で身につけたものである。

こうした音声の質、すなわち声の高さ、大きさ、響き方などを客観的に捉える指標を、「音響特徴量」と呼ぶ(図 1)。これには音程を示す基本周波数、音量、発話スピードや発話頻度などの実際の聴こえ方を数値化したものが含まれる。音響工学の領域では、直接的な音の聴こえ方を評価するのではなく、音声を物理学的な波として分析した音響特徴量が、音声認識に応用されている。音響特徴量の例を表 1 に示す。

主な音響特徴量には、基本周波数、フォルマント、高調波対ノイズ比、ジッター、シマー、ゼロクロス率線形予測符号(LPC)、知覚的線形予測(PLP)、メル周波数ケプストラム係数(MFCC)があげられる(Eyben et al., 2010)。このうち、基本周波数は音声スペクトルにおいて最も低い周波数ピークであり、一般に音程として認識される。フォルマントは音声スペクトルにおいて基本周波数より高い複数の周波数のピークであり、音色や音質が影響する声道の共振周波数である。高調波対ノイズ比とは、音声の付加ノイズの量を反映するものであり、音声の濁りが影響すると考えられている。ジッターとは時間軸方向での揺らぎを示し、シマーとは振幅方向での揺らぎを示す。ゼロクロス率とは音声波が時間軸を通過する頻度であり、音程の影響を受けるが音声の明瞭さも反映する。フォルマントを予測し、その効果を信号から除去し、残った部分の強さと周波数を予測することで話声信号を分析する線形予測符号(LPC)やLPCを人間の聴覚特性に合わせた知覚的線形予測(PLP)は音声圧縮、音声認識に用いられる。加えて、音声認識には、音声スペクトルを複数の次元で分解しスペクトル成分を概形化するメル周波数ケプストラム係数(MFCC)も用いられる。

### 4. MFCC

MFCCはMermelstein(Davis and Mermelstein, 1980)によって1970年代に導入され、それ以降声道の変化を反映する音響特徴量(Yinghua Zhu et al., 2013)として、スマートフォンなどの音声認識に幅広く応用されてきた。MFCCの導出過程は、1)周波数から高速フーリエ変換(FFT)スペクトルを算出する、2)人間の聴覚特性に基づくメル尺度上に配置されたフィルタバンクの

出力を抽出する、3)離散コサイン変換からケプストラム係数を得る、である(Sahidullah and Saha, 2012)。1)により周波数特性を得ることができ、2)により人間が認識しやすい周波数領域において周波数特性を平滑化し概形を得る。さらに 3)によりその周波数特性の変化を複数の成分に分割することができる。Mitrovićらによると、MFCC 0 がスペクトルの平均パワーを示し、MFCC 1 がスペクトルの大まかな形を示し、MFCC 2 を含むそれより高次の係数がより細かいスペクトルの形を示すとされる(Mitrović et al., 2010)。そのため、より低次の係数では個人差の大きい周波数成分を除外することができ、MFCC 0 以外は音量の影響を除外することができる。我々は、音声とストレスの関連を検討する目的で、健常者 106 名を対象として、ストレス課題前後の音声を録音し、課題実施群と未実施群において MFCC やゼロクロス率、基本周波数、高調波対ノイズ比などの音響特徴量を比較し、MFCC に関連する音響特徴量で差を認めたことを報告した(田口 et al., 2014)。

## 5. 音声と抑うつ状態に関する報告

音声と抑うつ状態との関連については、これまで様々な報告がされてきた。1980 年代には Streeter (Streeter et al., 1983)らや、Tolkmit ら(Tolkmitt and Scherer, 1986)が、感情ストレスにより基本周波数が減少したと報告した。1990 年代には Flint ら(Flint et al., 1992)が、パーキンソン病患者と大うつ病性障害患者の音声を比較し、後者において発話頻度が減少した、と報告した。Nilsonne らはうつ病患者において基本周波数が低下し、質問に対する返答までのポーズ時間が延長すると報告した(Nilsonne et al., 1988)。また、Cannizzaro らは、ハミルトンうつ評価尺度と発話頻度、ピッチの変化、発話の停止時間の割合が関連すると報告した(Cannizzaro et al., 2004)。いくつかの報告は音声が発療効果判定に使用できるかを検討しており、Mundt らはうつ病の治療反応性とピッチの変化、発話中の停止時間、発話のスピードとの関連を報告している(Mundt et al., 2012)。また、Alpert らはコントロールと比較して、うつ病患者で発話の抑揚が減少しており、治療が抑揚に影響することを報告した(Alpert et al., 2001)。しかし、発話頻度や停止時間を検出するためには比較的長い録音時間を要し、抑揚や基本周波数は個人差が大きいため、個人間の比較には使用しづらかったという。

これらの報告に加え、近年では実際の聴こえ方とは直接関連しない様々な種類の音響特徴量を解析する技術が進んでいる。特に音響工学分野では、INTERSPEECH 2009 Emotion Challenge(Schuller et al., 2009)や INTERSPEECH 2010 Paralinguistic Challenge(Schuller et al., 2010)を始めとした取り組みにおいて、様々な音響特徴量を用いて、感情やパラ言語情報を検出する報告がされている。INTERSPEECH 2009 Emotion Challenge では、音声から感情を認識するための標準化された方法を模索する目的で、あらかじめ決まった音声ライブラリを用い、

OpenSMILE というソフトウェアにより、特定の音響特徴量を抽出して感情認識を行う取り組みがされた。INTERSPEECH 2010 Paralinguistic Challenge では、同様の方法論で、音声から年齢、性別、話者の興味の度合いについて検出する取り組みがされた。MFCC を含めた音響特徴量と抑うつとの関連も報告されているが、これらの報告は多くが工学系研究者による幅広い音響特徴量を総合して自動的にうつ病と健常者を判別する機械学習の研究である (Akkaralaertsest and Yingthawornsuk, 2015; Cummins et al., 2015b; Solomon et al., 2015; Teintipsiri and Yingthawornsuk, 2014)。MFCC などの音響特徴量と抑うつとの関連をその病態生理も含めて検討した文献は、これまでにほとんどない。

## 6. 本研究の目的

音声をうつ病の診断（躁うつ病との鑑別やうつ病の分類）、ならびに症状評価のバイオマーカーに用いることができれば、スクリーニング、鑑別診断、症状経過などを客観的に評価する方法として有用となる。そこで今回我々は、音響特徴量をうつ病のバイオマーカーとする適用可能性を検討することを研究目的として、以下の研究を実施した。

まず研究 1 として、有用な音響特徴量の候補を挙げるべく、うつ病患者と健常者において音響特徴量が異なるか、さらにその音響特徴量を用いてうつ病患者を判別することができるかを検討した。

次に研究 2 として、治療による抑うつ状態の改善により音響特徴量に変化するか、電気痙攣療法の治療前後において抑うつ状態の変化と音響特徴量の相関を検討した。

最後に研究 3 として、うつ病患者と統合失調症患者、健常者の 3 群を比較し、うつ病と統合失調症の鑑別に音声を用いることができるか、を検討した。

## 研究 1 音響特徴量によるうつ病患者と健常者の判別

## 1. 目的

音響特徴量をうつ病の診断・症状評価のバイオマーカーにする適用可能性を検討することを目的に、音響特徴量が、うつ病患者とコントロールを判別することができるかどうか、抑うつ症状と相関するかどうかを検討する。

## 2. 対象と方法

### 2.1. 対象

うつ病群は、2013年11月から2015年3月の期間に、筑波大学附属病院精神神経科、筑波大学保健管理センター、茨城県立こころの医療センター、栗田病院に通院するうつ病患者のうち、研究同意が得られた38名を対象とした。うつ病群の全ての対象者は、精神科医により精神疾患の診断・統計マニュアル第5版(DSM-5)に基づく臨床面接を受け、うつ病性障害の診断基準を満たした。音声適切に録音されていなかった2名のうつ病群対象者を除外した。

コントロール群は、2013年11月から2015年3月の期間に、筑波大学附属病院の精神神経科の職員、研修医、実習中の学生、茨城町役場職員のうち、研究同意が得られた66名の一般健康人を対象とした。このうち、音声が正しく記録されていなかった1名、精神疾患や重症の頭部外傷の既往に該当する2名を除外した。さらに、Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version : QIDS-SRJ (Fujisawa, 2010; Rush et al., 2003)のスコアがカットオフポイントである6点以上の27名を除外した。

結果として、36名(男性22名、女性14名、年齢幅21-79歳、平均年齢 $44 \pm 16.3$ (標準偏差)歳)のうつ病性障害患者によるうつ病群と、36名(男性16名、女性20名、年齢幅22-58歳、平均年齢 $38 \pm 10.4$ 歳)のコントロール群が研究対象となった。

うつ病群において、罹病期間の平均は $7.00 \pm 5.73$ (平均 $\pm$ 標準偏差)年、中央値は5.47年であった。1名を除いて何らかの処方を受けており、服用中の向精神薬の等価換算値は、抗うつ薬がイミプラミン換算で $111.04 \pm 72.4$ mg、抗精神病薬がクロルプロマジン換算で $54.1 \pm 5.73$ mg、抗不安薬及び睡眠薬が $9.87 \pm 11.0$ mgであった(Inada and Inagaki, 2015)。すべての対象者は日本人である。対象について表2に示す。

本研究は筑波大学医の倫理委員会(814,814-1)、筑波大学附属病院倫理委員会(H26-7)、茨城県立こころの医療センター倫理委員会(受付番号30,49,60)、栗田病院倫理委員会(第043号)の承認を受けている。

## 2.2 音声の録音

音声はタブレット端末 Google Nexus 7 (TM) に、マイクロフォンと口との距離が大きく異なることを避けるために受話器型ヘッドセット (Hashy SP-2411) を接続し録音した (図 2)。受話器型ヘッドセットに内蔵されたマイクロフォンの周波数特性を図 3 に示す。このヘッドセットは一般使用を目的とした汎用品であり、その周波数特性は中低域の感度がやや高いものである。音声録音のガイダンス用アプリケーションを開発してタブレット端末にインストールし、被験者は音声と画面のガイダンスに従ってこれらのタスクを行った。被験者は、タブレット端末の画面、「枠の中に書かれている 012-345-6789 という数字を読み上げてください」という音声と合図の beep オンの後に、「012-345-6789」(ゼロイチニ、サンヨンゴ、ロクナナハチキュウ)を読み上げた。臨床応用可能性を考慮し短いフレーズとし、感情による測定バイアスを少なくするために数字を採用し、電話番号のように読み上げるタスクとした。次に、うつ病で機能が低下しているとされている前頭前野の機能を反映し、近赤外線スペクトロスコーピー (NIRS) でも採用されている言語流暢性課題(Masuda et al., 2017)を行った。「『あ』から始まる単語をなるべくたくさん挙げてください。制限時間は 30 秒です。それでは始めます」という音声と合図の beep 音の後、30 秒間で「あ」、「う」、「お」、それぞれから始まる単語をなるべくたくさん挙げるタスクを行った。その後、再度音声と画面の案内に合わせて「012-345-6789」と読み上げた。これらのタスクは全て日本語で行った。音声は全て、16 bit PCM, 22.05kHz で記録し、機器の録音音量、案内音声の音量を統一した。このため、記録した音声のノーマライズは行わなかった。

## 2.3. 音響特徴量の抽出

音響特徴量の抽出に先立ち、発声区間の前後の無発声区間を手動で削除した。音響特徴量の抽出には OpenSMILE v2.1.0(Eyben et al., 2010)を用い、設定パラメータとして”Feature set of INTERSPEECH 2009 Emotion Challenge”プリセットを使用した。このプリセットは INTERSPEECH 2009 conference において音声から感情を識別する取り組みに用いられたものである (Schuller et al., 2009)。音響特徴量として、音量の二乗平均平方根、12次元の MFCC、ゼロクロス率、高調波対ノイズ比、基本周波数が抽出され、これらに関する音響特徴量のうち、発声区間の平均値について検討した。音声ファイル (ビット深度 16bit、サンプリング周波数 22.05kHz の WAVE ファイル) は OpenSMILE に入力され、10ms 毎に 25ms のフレームが生成され、解析された。

各音響特徴量について概説すると、音量の二乗平均平方根 (RMS) は時間と共に変化する

音量を実効的に示す指標であり、 $x$ を各フレームにおける音量、 $N$ をフレーム数として、

$$\text{RMS}[x] = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i)^2}$$

と示される。

MFCC の簡単な導出過程と意義は研究背景で示した通りである。実際の手順としては、まず、高速フーリエ変換を行う準備として、Hanning 窓を掛け区間外のデータを 0 とし

(windowing)、対象となる音声の後に 0 の要素を加え (zero padding)、離散コサイン変換を行う。N個のデータからなる信号を $f(x)$ 、変換後の信号を $a_n$ 、 $b_n$ とすると、

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$$

となり、得られた振幅

$$|f(x)| = \sqrt{a_n^2 + b_n^2}$$

をメル尺度による三角窓を持つフィルタバンクに通し、対数をとる。

メル尺度は、

$$\text{Mel}(f) = 1127.01048 \log \left( 1 + \frac{f}{700} \right)$$

で計算される。

0Hz から 8000Hz の間において、このメル尺度上で等間隔となる 26 点を得る。この点は単位を Hz とすると、68.48、143.66、226.19、316.80、416.27、525.47、645.35、776.97、921.46、1080.08、1254.22、1445.40、1655.27、1885.69、2138.64、2416.33、2721.20、3055.88、3423.31、3826.69、4269.52、4755.68、5289.39、5875.32、6518.57、7224.74 となる。この 26 点を中心とし、0Hz、8000Hz と各中心点 26 点の計 28 点において、中心点と隣り合う 2 点と中心点とを結ぶ 26 個の三角形のフィルタをメルフィルタバンクと言い、高速フーリエ変換で得られた振幅をメルフィルタバンクに通し、対数をとったものを FBANK と呼ぶ。

これに離散コサイン変換を掛け、離散的な N 次 cos を基底とする一次結合の係数に変換する。

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left\{ \frac{\pi}{N} \left( n + \frac{1}{2} \right) k \right\}$$

この係数をメル周波数ケプストラム係数という。フィルタバンクの数を $N$ とし、 $m_j$ を対数フィルタバンクの振幅とすると、係数 $c_i$ は、



$$C_i = \sqrt{\frac{2}{N}} \sum_{j=1}^N m_j \cos\left\{\frac{\pi i}{N}(j - 0.5)\right\}$$

と示される。

ゼロクロス率 (zcr) は、解析区間において波の正負が変化する 0 を通過する頻度であり、解析区間  $T$  における音声信号  $s$  として、

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{\mathbb{R} < 0}(s_t s_{t-1})$$

と示される。

高調波対ノイズ比 (HNR) は音声に対する付加ノイズの相対的な量であり、高調波を  $H$ 、付加ノイズを  $N$  として、

$$\text{HNR} = 10 \times \log_{10}\left(\frac{H}{N}\right)$$

と示される。

基本周波数は最も低い周波数ピークであり、一般に音程と捉えられる。

なお、音響特徴量の再現性については、事前に異なる環境で同一人物の音声を録音し、高い再現性が得られること、音量の違いによって影響を受けないことを確認した。

## 2.4. 統計解析

初めに、基本属性別、うつ病群とコントロール群の音響特徴量の違いについて Student の  $t$  検定による群間比較を行った。うつ病とコントロールの判別因子となりうる指標については、ステップワイズ法を用いた判別分析、ならびに Receiver Operating Characteristic (ROC) 解析を行い、曲線下面積 Area under curve (AUC)、感度、特異度を算出した。次に、うつ病群、コントロール群それぞれにおいて QIDS-SRJ のスコアと音響特徴量との相関解析を行った。さらに年齢及び性差の影響を鑑別するために、年齢、性別、群 (うつ病群、およびコントロール群) を説明変数とし、判別分析で有用であった音響特徴量を目的変数とした重回帰分析を行なった。最後に、うつ病群とコントロール群の判別に有用であった音響特徴量について、結果の再現性を検討するため、反復測定の一元配置分散分析を行い、言語流暢性課題の前後の音声を比較した。なお、反復測定分散分析においては、有意であった場合には post-hoc 解析を行うのが通常であるが、統計モデルがうつ病群、コントロール群の 2 群に対して、課題の前後の音声の変化を検討するというものであったため、post-hoc 解析は行わなかった。

統計解析には IBM SPSS Statistics 22 (IBM corp, Armonk, NY, USA) を用いた。

### 3. 結果

#### 3.1. 各群の音響特徴量の差の検討

うつ病群とコントロール群を比較し、Bonferroni の多重比較補正後も有意な差を認めたのは MFCC 2 であり、うつ病群がコントロール群に比して有意に高値であった ( $t=-6.728$ ,  $p=5.33\times 10^{-9}$ , Cohen's  $d=-1.72$ )。その他の音響特徴量では有意な差は認めなかった。Shapiro-Wilks normality test で、MFCC 2 は、全対象の  $W = 0.98$ ,  $p\text{-value} = 0.30$ 、うつ病群の  $W = 0.97$ ,  $p\text{-value} = 0.42$ 、コントロール群の  $W = 0.96$ ,  $p\text{-value} = 0.3028$  と、いずれも正規性が確認された。

男女別での検討でもそれぞれで MFCC 2 のみでうつ病群が有意に高かった (男性:  $t=-4.903$ ,  $p<0.001$ , Cohen's  $d=-1.63$ ; 女性:  $t=-5.122$ ,  $p<0.001$ , Cohen's  $d=-1.81$ )。結果は表 3 に示した。

40 歳以下 (younger group) とそれ以上 (older group) とで分けたところ、younger group はうつ病群 17 名 (男性 12 名、女性 5 名)、コントロール群 20 名 (男性 12 名、女性 8 名) となり、older group はうつ病群 19 名 (男性 10 名、女性 9 名)、コントロール群 (男性 4 名、女性 12 名) となった。各グループで同様に差を比較したところ、いずれも MFCC 2 のみうつ病群が有意に高値であった (younger group:  $t=-5.327$ ,  $p<0.001$ , Cohen's  $d=-1.80$ ; older group:  $t=-4.858$ ,  $p<0.001$ , Cohen's  $d=-1.69$ )。

#### 3.2. MFCC による判別分析

ステップワイズ法による判別分析を行った結果、MFCC 2 のみが判別に採用され、うつ病群とコントロール群を、感度 77.8%、特異度 86.1%、判別確率 81.9% で判別できた (Wilks'  $\lambda = 0.567$ ;  $\chi^2 = 39.456$ ,  $df = 1$ ,  $p < 0.001$ )。結果は表 4 に示した。図 4 は、各群における MFCC 2 と判別された群についての MFCC 2 の分布である。うつ病群のうち向精神薬を内服していない対象者は、うつ病群に判別された。また、MFCC 2 の ROC 解析の結果、AUC は 0.88 であり、ROC 曲線において左上隅に最も近づく値を閾値としたところ、感度 80.6%、特異度 83.3% であった (図 5)。

#### 3.3. 各群における周波数スペクトルの差の検討

MFCC 2 において差を認めたという結果について、両群の音声スペクトルを比較した。図 6 は MFCC を計算する過程にある、メル尺度に基づくフィルタバンクを通し平滑化した周波数特性 (FBANK) を示す。コントロール群と比較し、うつ病群では 2000Hz から 3000Hz を中心とし

た周波数帯域においてエネルギーが低く、Bonferroni 補正後も、有意な群間差を認めた。図 7 はコントロール群とうつ病群それぞれの平均値の差分と、MFCC 0 から 3 までの離散コサイン変換における基底を示す。差分のパターンは MFCC 2 の基底に近い形態であり、これが MFCC 2 の変化とうつ病群における音声エネルギーの低下パターンとの関連を説明しうると考えられた。

#### 3.4. 音響特徴量と QIDS-SRJ の相関解析

Bonferroni 補正後、抑うつ尺度 QIDS-SRJ と有意な相関を示す音響特徴量はなかった。MFCC 2 と QIDS-SRJ においても、うつ病群の  $r = -0.0233$ ,  $p\text{-value} = 0.893$ 、コントロール群の  $r = 0.254$ ,  $p\text{-value} = 0.136$  と有意な相関を示さず、MFCC 2 と QIDS-SRJ の下位項目（睡眠、抑うつ、食欲・体重、集中力・決断、自己評価、希死念慮、興味、疲労、精神運動状態）においても有意な相関は認めなかった。

#### 3.5. MFCC 2 における重回帰分析

年齢、性別、うつ病の有無（うつ病群、コントロール群）を説明変数とし、MFCC 2 を目的変数とする重回帰分析を実施したところうつ病の有無の  $\beta = 0.637$  ( $p\text{-value} < 0.001$ )、年齢の  $\beta = 0.068$  ( $p\text{-value} = 0.482$ )、性別の  $\beta = -0.039$  ( $p\text{-value} = 0.688$ ) であり、うつ病の有無のみが有意に MFCC2 の違いを説明した。 $R^2$  は 0.438 と中程度のモデル説明力を示した。MFCC 2 はうつ病の有無に影響を受けるが、年齢、性別からはほとんど影響を受けなかった。結果を表 5 に示す。

#### 3.6. 言語流暢性課題前後の音声における反復測定の一元配置分散分析

MFCC 2 の結果の再現性を検討するため、言語流暢性課題前後の MFCC 2 について反復測定一元配置分散分析を行なった。言語流暢性課題と群の有意な交互作用は認めなかった ( $F=3.511$ )。一方、全対象者における言語流暢性課題の主効果は統計学的に有意 ( $F=4.508$ ,  $p<0.05$ ) であり、群による主効果もまた有意 ( $F=51.117$ ,  $p<0.001$ ) であった。図 8 に言語流暢性課題の前後における各群の MFCC 2 の平均値を示した。

## 4. 考察

本研究において、MFCC 2 の値がうつ病群においてコントロール群と比較し、有意に高いことがわかった。この MFCC 2 の違いは年齢や性別の影響を受けなかった。MFCC2 は高い感度と特異度でうつ病とコントロールを判別できた。一方、MFCC 2 はうつ病の重症度とは相関せず、言語流暢性課題の影響を受けなかった。

### 4.1. MFCC 2 の変化はなぜ起こるのか

これまで MFCC とうつ病の関連については、いくつかの研究報告があるが(Akkaralaertsest and Yingthawornsuk, 2015; Cummins et al., 2015a; Solomon et al., 2015; Teintipsiri and Yingthawornsuk, 2014)、MFCC 2 を含む MFCC のそれぞれの次元における関連報告はない。さらに MFCC とうつ病との関連がなぜ生じるか考察はされていない。

本研究では、音声スペクトルを検討した結果、MFCC 2 の違いがうつ病患者において 2000Hz から 3000Hz の周波数帯域でエネルギーが低いことと対応することがわかった。一般的に、この周波数帯域は発声の広がり方に影響するとされており(Bele, 2006; Leino et al., 2011; Sundberg, 2001)、2000Hz から 3000Hz の周波数帯域は人間の聴感上の音量が等しくなるよう設定された等ラウドネス曲線 (ISO 226:2003) において、最も感度が高い帯域である。このような特定の周波数帯域でエネルギーの低下が認められたが、音量では有意な差がなかったことから、今回の結果は、うつ病患者の音量の差よりも、音声がかもって聞こえることと関連することを示唆している。

MFCC 2 のような低次元の MFCC は声帯振動の影響を受けないとされており、声道の変化を反映するものと考えられる。声道は、声帯から口唇まで、咽頭、喉頭、鼻腔、口腔といった構造物によって構成され、歯や筋緊張、舌や下顎の動きなどにより音質が変化する。Yinghua らは MRI を用いて MFCC と声道が関連すると報告している(Yinghua Zhu et al., 2013)。このことから、うつ病においては声道の変化が生じており、これによって声がかもり、音響特徴量のうち MFCC 2 が変化したと推測される。

以上の考察から、MFCC 2 はうつ病の有力なバイオマーカーとなりうることが示唆された。

### 4.2. 基本周波数

多くの研究で、基本周波数と、抑うつ状態を含む精神状態との関連が報告されているが、今回の結果からはうつ病群とコントロール群で有意な差を認めなかった。基本周波数は音程を反映

し、個人差の大きい特徴量である。今回は音声サンプル全体の基本周波数の平均値を用いて解析を行なったが、「012-345-6789」という音声には、読み上げの途中で発声していない区間が存在し、基本周波数の数値に影響を与え、実際の基本周波数とは異なる結果となっていた可能性がある。今後の研究において基本周波数がうつ病を識別するのに有用であるかどうかは、読み上げ課題を変えた検討が必要である。

#### 4.3. 小括

本研究ではうつ病患者の音声において音響特徴量である MFCC 2 が変化し、その変化が 2000Hz から 3000Hz の音声周波数帯域における声道特性の変化と関連することを示した。この MFCC 2 の変化はうつ病患者の声がこもって聞こえるという臨床的な印象と一致する。MFCC 2 はうつ病の診断において有用なバイオマーカーとなる可能性が示唆された。

## 研究 2 治療による抑うつ症状と音響特徴量の変化

## 1. 目的

研究 1 では MFCC 2 を用いてうつ病患者の判別ができる可能性が示唆されたが、横断面的には症状との相関は認められず、症状評価への適用可能性の吟味は不十分であった。電気けいれん療法 (ECT) は、短期に症状改善の効果がある治療法であり、治療前後で比較的短期に症状変化の比較が可能である。このため、ECT 前後の音声と抑うつ症状を評価し、治療前後で症状の改善と音声の変化が相関するか、治療前後の比較検討を行う。

## 2. 対象と方法

### 2.1. 対象

筑波大学附属病院精神神経科で ECT を行う予定のうつ病患者のうち、研究について説明をし、同意が得られた 10 名 (男性 5 名、女性 5 名、年齢幅 27-77 歳、平均年齢  $55 \pm 19.0$  (標準偏差) 歳) を対象とした。全ての対象者は、精神疾患の診断・統計マニュアル第 5 版 (DSM-5) に基づいた臨床面接を受け、うつ病性障害の診断基準を満たした。罹病期間の平均は  $4.9 \pm 4.8$  (標準偏差) 年であり、中央値は 14 年であった。教育年数の平均は  $13.30 \pm 2.50$  (標準偏差) 年であり、中央値は 2.5 年であった。使用された向精神薬の等価換算値 (平均  $\pm$  標準偏差) は、抗うつ薬がイミプラミン換算で  $134.5 \pm 188.8$  mg、抗精神病薬がクロルプロマジン換算で  $123.0 \pm 120.0$  mg、抗不安薬及び睡眠薬がジアゼパム換算で  $5.12 \pm 4.65$  mg であった。対象の属性は表 6 に示した。

本研究は筑波大学附属病院、医の倫理委員会の承認を受けている。

### 2.2. 抑うつ症状の評価

ECT 前後における抑うつ症状の評価として、Quick Inventory of Depressive Symptomatology, Japanese version (QIDS-J) (藤沢, 2010; Rush et al., 2003) を用い、ECT を実施する前日と終了翌日に実施した。

### 2.3. 音声の録音・抽出

音声の録音・抽出については研究 1 の方法に準じて実施した。録音は、同一患者について ECT を実施する前日と終了翌日に行なった。

## 2.5. 統計解析

ECT 前後の QIDS-J 総スコア、下位項目（睡眠、抑うつ、食欲・体重、集中力・決断、自己評価、希死念慮、興味、疲労、精神運動状態）のスコア、および音響特徴量に対し、対応のあるサンプルの  $t$  検定を行なった。

また、ECT 前後の QIDS スコアの差分（「ECT 後の QIDS-J スコア」-「ECT 前の QIDS-J スコア」）と ECT 前後の音響特徴量の差分（「ECT 後の音響特徴量」-「ECT 前の音響特徴量」）について、Pearson の相関解析を行なった。

全ての統計解析は IBM SPSS Statistics 22 (IBM corp, Armonk, NY, USA) を用いた。

## 3. 結果

### 3.1. ECT 前後における QIDS-J スコアの対応のあるサンプルの $t$ 検定と相関解析

ECT 前後の QIDS-J スコアおよび下位項目スコアにおける、対応のあるサンプルの  $t$  検定では、Bonferroni の補正後、総スコアと下位項目のうち集中力・決断においてのみ  $p < 0.05$  の有意差を認めた。結果は表 7 に示した。QIDS-J の総スコアおよび下位項目のスコアそれぞれにおける ECT 前後の相関解析では有意な相関を認めた項目はなかった。

### 3.2. ECT 前後における音響特徴量の相関解析と対応のあるサンプルの $t$ 検定

ECT 前後の音響特徴量についての Pearson の相関解析では、音量の二乗平均平方根、MFCC 4、MFCC 7、MFCC 10、MFCC 11、高調波対ノイズ比、基本周波数で有意な相関が得られ、特に高調波対ノイズ比および基本周波数で強い相関（高調波対ノイズ比： $r=0.944$ 、 $p < 0.0001$ 、 $\text{intercept}=0$ 、 $\text{slope}=1.001$ 、基本周波数： $r=0.930$ 、 $p < 0.0001$ 、 $\text{intercept}=0$ 、 $\text{slope}=1.006$ ）が得られた。また、ECT 前後の音響特徴量において対応のあるサンプルの  $t$  検定を行ったところ、音量の二乗平均平方根、12 次元の MFCC、ゼロクロス率、高調波対ノイズ比、基本周波数のうち、ECT 前後において有意な差を認めた音響特徴量はなかった。

### 3.3. ECT 前後における QIDS-J スコアの変化と音響特徴量の変化の相関解析

ECT 前後の QIDS-J 総スコア、下位項目の差分は、ECT 前後の音響特徴量との相関解析にお



いて、Bonferroni の補正後、有意な相関を認めた項目はなかった。

## 4. 考察

### 4.1. ECT 前後の抑うつ症状の変化

ECT 前後における QIDS-J スコアおよび下位項目スコアの差の検討では、総スコアと下位項目の集中力・決断において  $p < 0.01$  の有意差を認めたことから、ECT により抑うつ症状が改善し、特に集中力・決断での改善が早期に現れやすいと考えられた。しかし、それぞれの項目は ECT 前後で有意な相関を認めず、個人により ECT で改善した抑うつ症状の種類にばらつきがあることが考えられた。

### 4.2. ECT 前後の音響特徴量の変化

ECT 前後の音響特徴量について Pearson の相関解析を行い、結果で示した複数の音響特徴量に有意な相関が認められた。特に高調波対ノイズ比と基本周波数に強い相関を認め、切片を 0 とした時に線形曲線の傾きも 1 に近かったため、これらの音響特徴量は ECT 実施による変化が少なく、個人差を反映しやすいものと考えられた。

また、対応のあるサンプルの  $t$  検定では、統計学的に有意な差のある音響特徴量は認めず、ECT による抑うつ症状の改善において、研究 1 で認めた、うつ病患者で MFCC 2 が上昇しているという結果の関連は見いだせなかった。また、今回の MFCC 2 の平均は ECT 前で -10.2、ECT 後で -10.0 であったが、研究 1 ではコントロール群の MFCC 2 の平均が -14.3 に対し、うつ病群の平均が -10.2 であり、ECT 前後共に研究 1 のうつ病群の結果とほぼ同様であった。QIDS-J 上、ECT による抑うつ症状の改善は認めたものの、下位項目は抑うつ、希死念慮、食欲・体重で有意差を認めなかったことから、短期的な改善では MFCC 2 が変化しないことが示唆された。

ECT は重度のうつ病に対して効果的な治療であるとされ、“Depression in adults: recognition and management | Guidance and guidelines | NICE,” 2009)、週 2 回の ECT 継続期間において、ECT 3 回目で半数以上に抑うつ改善の反応を認め、ECT 6 回目で 34%が寛解となり、ECT 10 回目には 65%が寛解となることも報告されている (Husain et al., 2004)。しかし、抑うつ症状の改善は早期に認められるものの、身体的な変化の有無については明らかでない。Abbott によると、ECT 前と ECT 後 5 日以降で resting-state functional magnetic resonance imaging (resting-state fMRI) を行なったところ、海馬領域の機能的結合性の改善、および海馬の容積の増加と抑

うつ症状の減少とが相関しており (Abbott et al., 2014)、ECT による抑うつの改善と海馬、扁桃体の変化との関連が報告されてきた (Joshi et al., 2016; Tendolkar et al., 2013)。しかし、MFCC 2 の変化は、声道の変化を反映しているものと考えられ、その変化が生じるまでには心理尺度の改善や脳の変化よりも長期間を要すると考えられる。MFCC 2 が抑うつの改善に伴い変化するかどうかを結論づけるためには、寛解後、より長期のフォローアップにおける音声の比較が必要である。

#### 4.3. ECT 前後での QIDS-J スコアの変化と音響特徴量の変化の関係

ECT 前後で変化した QIDS-J スコアと今回検討した音響特徴量との間で有意な相関が得られなかったが、QIDS-J は ECT 後に改善しており、今回検討した音響特徴量は MFCC 2 を含め、抑うつ症状の改善と同時に変化しないことが考えられた。前述のように、身体的な変化には ECT 期間よりも長期間を要すると仮定すれば、より長期のフォローアップにおいては QIDS-J とも関連がみられる可能性もあり、さらなる検討を要する。

#### 4.4. 小括

ECT 前後の音声と QIDS-J スコアを比較した。QIDS-J は ECT 前後で改善したが、今回検討した音響特徴量では変化を認めず、QIDS-J との相関も認めなかった。短期的な抑うつ症状の改善では、声道の身体的な変化にまでは至らないことが示唆された。

### 研究3 うつ病患者、統合失調症患者、健常者の音声の比較

## 1. 目的

研究1ではうつ病患者と健常者の比較を行ったが、MFCC2がバイオマーカーとなるためには、うつ病と他疾患による抑うつ状態との鑑別ができることが期待される。そこで研究3では、研究1の結果に統合失調症患者の音声を加えて比較し、統合失調症患者における抑うつの有無、MFCCの変化、鑑別に有用な音響特徴量の有無について検討を行った。

## 2. 対象と方法

### 2.1. 対象

うつ病群、統合失調症群として、2013年11月～2016年12月の期間に、筑波大学附属病院精神神経科、筑波大学保健管理センター、茨城県立こころの医療センター、栗田病院に通院するうつ病患者、統合失調症患者のうち研究について説明をし、同意が得られたうつ病患者48名、統合失調症患者26名を対象とした。うつ病群、統合失調症群の全ての対象者は、精神疾患の診断・統計マニュアル第5版(DSM-5)に基づいた臨床面接を受け、うつ病性障害あるいは統合失調症の診断基準を満たした。音声が適切に録音されていなかったうつ病群2名、統合失調症群2名の対象者を除外し、うつ病群46名、統合失調症群24名を採用した。コントロール群は、2013年11月～2016年12月の期間に、筑波大学附属病院の精神神経科の職員、研修医、実習中の学生、茨城町役場職員のうち、過去や現在に精神疾患の既往がない66名の一般人から成り、音声が正しく記録されていなかった1名、精神疾患の既往や重症の頭部外傷に該当する2名を除外した。さらに、Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version (QIDS-SRJ) (藤沢, 2010; Rush et al., 2003) を実施し、そのスコアがカットオフポイントである6点より高い27名を除外した。また、統合失調症群には日本語版簡易精神症状評価尺度(BPRS)を実施した。結果として、45名(男性27名、女性18名、年齢幅21-79歳、平均年齢 $44 \pm 16.1$  (標準偏差) 歳)のうつ病群と、24名(男性9名、女性15名、年齢幅19-61歳、平均年齢 $40 \pm 12.4$  (標準偏差) 歳)の統合失調症群、39名(男性16名、女性23名、年齢幅22-58歳、平均年齢 $37 \pm 10.6$  (標準偏差) 歳)のコントロール群が構成された。男女比の差についてはカイ2乗検定で有意ではなかった。うつ病群において、罹病期間の平均は $8.4 \pm 7.5$  (標準偏差) 年、中央値は5.5年であった。統合失調症群において、罹病期間の平均は $14.3 \pm 13.0$  (標準偏差) 年、中央値は8.0年であった。うつ病群で使用された向精神薬の等価換算値(平均 $\pm$ 標準偏差)は、抗うつ薬がイミプラミン換算で $122.0 \pm 87.7$  mg、抗精神病薬がクロロプロマジン換算で $38.5 \pm$

107.0mg、抗不安薬及び睡眠薬が  $10.0 \pm 10.0$  mg であった。統合失調症群で使用された向精神薬の等価換算値（平均±標準偏差）は、抗うつ薬がイミプラミン換算で  $20.1 \pm 68.5$  mg、抗精神病薬がクロルプロマジン換算で  $669.1 \pm 493.9$ mg、抗不安薬及び睡眠薬が  $11.1 \pm 13.9$  mg であった。すべての対象者は日本人であり、うつ病群、健常者群については研究 1 の対象者を全て含んでいる。対象について表 8 に示す。

## 2.2 音声の録音・抽出

音声の録音・抽出については研究 1 に準じて行った。

## 2.4. 統計解析

各群における音響特徴量の比較として、群を独立変数とし、音響特徴量を従属変数として一元配置分散分析を行なった。その後の検定には、Tukey の多重比較検定を用いた。

統合失調症群においては、音響特徴量と QIDS-SRJ および BPRS との相関解析を行なった。また、QIDS-SRJ のスコアが 11 以上（中等度以上の抑うつ）となる抑うつあり群と QIDS-SRJ が 11 未満となる抑うつなし群とに分け、Kolmogorov-Smirnov test による正規性の確認を行った上で MFCC 2 について各群の差の検定を行なった。

全ての統計解析は IBM SPSS Statistics 22 (IBM corp, Armonk, NY, USA) を用いた。

本研究は筑波大学附属病院倫理委員会、茨城県率こころの医療センター倫理委員会、栗田病院倫理委員会の承認を受けている。

## 3. 結果

### 3.1. 各群の音響特徴量の比較

Bonferroni の多重比較補正後、3 群間で有意差を認めたのは MFCC 2 のみであった (MFCC 2; F 値: 24.04、 $p=2.5 \times 10^{-9}$ 、コントロール群平均: -14.04、統合失調症群平均: -12.07、うつ病群平均: -10.41)。MFCC2 はうつ病群、統合失調症群、コントロール群の順に高かった。Tukey の二群間比較検定では、統合失調症群とコントロール群 ( $p<0.01$ )、うつ病群と統合失調症群 ( $p<0.05$ ) の間にそれぞれ有意差を認めた。各群における MFCC 2 の分布を図 9 に示す。

### 3.2. 音響特徴量と QIDS-SRJ、BPRS との相関

統合失調症群の MFCC 2 を含めた各音響特徴量において、Bonferroni の多重比較補正後、QIDS-SRJ、BPRS とともに有意な相関は認めなかった。

### 3.3 統合失調症群のうち抑うつが中等度以上ある群とない群の差の比較

統合失調症群（24 名）のうち中等度以上の抑うつと判断される QIDS-SRJ スコアが 11 以上となる 10 名（男性 2 名、女性 8 名）、QIDS-SRJ スコアが 11 未満となる 14 名（男性 7 名、女性 7 名）を対象とした。MFCC 2 においては Kolmogorov-Smirnov の正規性の検定で  $p=0.120$  であったため、Student- $t$  検定を行なった。MFCC 2 は  $t=-0.474$ 、 $p=0.640$  と抑うつの有無の 2 群間で、有意な差は得られなかった。

## 4. 考察

本研究では統合失調症患者を含めた 3 群で音響特徴量の比較を行った。統合失調症の MFCC 2 の値はうつ病患者と健常者の中間となり、健常者と統合失調症患者 ( $p<0.01$ )、統合失調症患者とうつ病患者 ( $p<0.05$ ) の間でそれぞれ有意差を認めた。しかし、QIDS-SRJ スコアの平均や標準偏差はうつ病患者、統合失調症患者でほぼ同様であった。このことから MFCC 2 の違いは統合失調症においても抑うつ状態を示す可能性があるが、うつ病よりは程度が弱いことが示唆された。また、うつ病患者、健常者においては QIDS-SRJ と各音響特徴量の相関は認めなかったが、統合失調症群においても QIDS-SRJ との相関はなく、中等度以上の抑うつ症状の有無によって群を分けて比較しても有意な差は認めなかった。また、主に統合失調症の陽性症状の程度を表す BPRS との相関も認められなかった。

### 4.1. 統合失調症の抑うつ症状

統合失調症における抑うつ症状は、陰性症状との区別も難しくその定義も様々であるが、自殺のリスクでもあるため重要視されている (Bartels and Drake, 1988)。統合失調症治療において、患者が病識を得ることが治療者の目標でもあるが、病識が高いことが抑うつ症状の存在と強く関連するという "insight paradox" も報告されており (Belvederi Murri et al., 2016)、必ずしも統合失調症の陽性症状や陰性症状と関連するとは限らない。このように統合失調症患者でも抑うつ症状は存在するが、その定義が様々であるため評価方法も定まったものがない。今回は

QIDS-SRJ スコアがうつ病患者と統合失調症患者でほぼ同様の結果であったが、QIDS-SRJ は自記式の評価尺度であり患者本人の主観に大きく影響されるため幻覚妄想状態の患者などは実際に抑うつ症状とは乖離した結果にもなりうることが考えられ、正確な抑うつの評価ができていない可能性がある。しかし、MFCC 2 の数値および QIDS-SRJ スコアともにコントロールとも異なる結果であったことから、統合失調症においても「声の通りが悪い」というような、うつ病に近い形での声道の変化を来しており、MFCC 2 などを用いて統合失調症の抑うつ症状を検出することができる可能性が示唆される。統合失調症患者の状態像は様々であるが、実臨床においてもうつ病患者のように声が聞き取りづらい統合失調症患者の症例を経験することがある。今後、統合失調症における抑うつ症状の評価方法の検討や対象者を増やすことが必要である。

#### 4.2. 小括

統合失調症患者の音声において MFCC 2 は健常者とうつ病患者の中間を示し、それぞれの群で有意な差を認めた。しかし、うつ病患者と同様、抑うつ症状との相関は認めず、陽性症状、陰性症状との相関も認めなかった。MFCC 2 の違いは統合失調症においても抑うつ状態を示す可能性があるが、うつ病よりは程度が弱いことが示唆された。

## 総合考察



## 1. 研究成果のまとめ

うつ病において音声が発断や鑑別に利用できるバイオマーカーとして用いることができるかを検討する目的で、本研究では、まず研究1として音響特徴量とうつ病との関連を検討し、次に研究2として治療による変化がないか検討し、最後に研究3として統合失調症との鑑別が可能か検討した。3つの研究成果を以下にまとめる。

研究1では、どのような音響特徴量とうつ病の検出に有用か、MFCCに着目して検討し、短い読み上げ音声におけるMFCC2の平均の値がうつ病患者と健常者で異なることを見出した。この知見は、人間の聴覚で最も敏感な周波数帯域における低下により音が聞き取りづらくなっている事と関連していると考察した。また、MFCC2はうつ病患者と健常者で差があるものの、抑うつの重症度との相関は認めず、年齢や性別による影響は受けにくく抑うつ状態の有無を反映する指標と考えられた。

研究2では、治療による抑うつの改善と音響特徴量の変化について検討を行い、短期的な改善ではMFCC2が変化しないことが考えられた。

研究3では、統合失調症とうつ病を鑑別する新たな音響特徴量を見出すことはできなかったが、統合失調症のMFCC2の平均がうつ病および健常者のMFCC2の平均の中間の値をとることがわかり、統合失調症における抑うつ症状を反映している可能性が考えられた。一方、統合失調症群において抑うつが中等度以上の群と未満の群においてMFCC2に有意な差は認めず、統合失調症群において抑うつ症状との関連は認められなかった。このことは、対象者が24名と少なく、患者が幻覚妄想状態から残遺状態まで様々な状態像を示していたことから正確な結果が得られなかったことも考えられた。

研究1から研究3を通して、短時間の読み上げ音声におけるMFCC2の発話区間平均値は、録音時の周辺環境や男女差、年齢差の影響を受けにくく、抑うつ状態の有無を反映する指標であると考えられ、長期的な治療で変化する可能性はあるが、短期間の改善では変化しないと考えられた。

## 2. 研究成果の解釈

音声は声帯振動が声道で共鳴することで形成されるが、声道は気道の一部であり、呼吸も音身に必要な要素である。今回は音声とうつとの関連を検討したが、呼吸とうつとの関連についてはこれまで多くの報告がされている。慢性閉塞性肺疾患(COPD)患者の40から60%に抑うつ症状があり(de Voogd et al., 2009)、COPD患者が酸素療法により抑うつ症状を呈したり、元々存在している抑うつ症状が増悪したりするとも報告されており、これらの背景に低換気によるCO<sub>2</sub>

濃度の上昇の関連も言及されている(Maurer et al., 2008)。また、抑うつ症状を示す閉塞性睡眠時無呼吸患者では、経鼻的持続陽圧呼吸 (CPAP) を行うことによって症状が改善した報告もある(Schwartz and Karatinos, 2007)。また、Sudarshan Kriya Yoga と呼ばれる呼吸リズムを整える運動を行うことにより抑うつが改善したという報告もある(Janakiramaiah et al., 2000)。

今回の研究において MFCC 2 が変化したことは、声道の変化により声がこもり、その聞き取りやすさが低下したと推察した。しかし前記のように気道が形態的に変化することと抑うつ症状との関連が指摘されている報告を踏まえると、うつ病による呼吸の変化が声の変化に関連した可能性も考えられる。

### 3. 音声はバイオマーカーになりえるのか

バイオマーカーとは、「通常の生物学的過程、病理学的過程、もしくは治療的介入に対する薬理的応答の指標として、客観的に測定され評価される特性」と定義されており、広義には日常診療で用いられるバイタルサインや、生化学検査、血液検査、腫瘍マーカーなどの各種臨床検査値や画像診断データなどが含まれる(Atkinson et al., 2001)。

バイオマーカーは、その目的に応じて診断マーカー、モニタリングマーカー、予後マーカー、予測マーカーなどが存在し、医療の質の向上に貢献している。研究のまとめで述べたように、我々の結果では、短時間の読み上げ音声において、MFCC 2 は抑うつ状態そのものを反映する指標であると考えられた。また、MFCC 2 はうつ病と統合失調症の音声において差を認めた。これらを考慮すると、音声、特に MFCC 2 はうつ病の診断マーカーになり得ると考えられる。

音声がうつ病の診断バイオマーカーになり得るとすると、うつ病の早期発見・スクリーニングに役立つ可能性がある。うつ病は、学業・仕事や日常生活に支障をきたしてはじめて周囲が気づくことが多く、早期発見は容易ではない。近年、本邦では平成 27 年 12 月よりストレスチェック制度(厚生労働省, 2015)が導入されたが、これは自記式検査であるため、職場関係を配慮して健康と答える恐れもあり、その診断的客観性には難がある。その点 MFCC2 のような音響特徴量は、客観的指標であるため、健診などでスクリーニングに用いれば、早期診断、早期介入が可能となる。また、児童・思春期もうつ病に対する早期介入が期待されている年代であるが、まだ早期診断に関するエビデンスが蓄積されていない(Cuijpers et al., 2006; Cullen et al., 2009)。介入方法の確立のためには診断バイオマーカーは必須であり、音声は一定の役割を果たしうる可能性がある。

これまで、数多くの企業で導入されている Nemesysco 社の Layered Voice Analysis(Nemesysco, 2007)や、国内でも日立システムズと PST 株式会社が協業した「音声こころ分析サービス」(株式会社日立システムズ, 2017)などが、音声により抑うつや感情を評価するサービスとして実用

化されているが、これらの技術については解析手法が不明であり、その手法の学術的なエビデンスの検討が課題と思われる。一方、MFCC2は音響工学分野では一般的な音響特徴量であり、一般公開されている無料のプログラムにより解析することが可能であり、その変化の理由についても周波数特性において声の聞き取りやすさと関連する声道の変化として説明が可能である。このため、より信頼性の高いバイオマーカーとなりうる。

一方、ECT患者の治療前後の音声では、うつ状態は改善したにもかかわらず、MFCC2には変化は認めなかった。このことから、短時間の音声から得られたMFCC2は、うつ病のモニタリングマーカーとはならない可能性がある。一方臨床現場では、ある程度話を聞いている間に、話のトーンや抑揚から、うつ状態を疑うことがある。そのことを考えると、数秒ではなく、数十秒の音声から得られる音響特徴量がモニタリングマーカーになる可能性もある。

#### 4. 研究の限界と今後の方向性

本研究には多くの限界がある。まず、サンプル数が少ないため、今後大規模サンプルで本研究の結果を確認する必要がある。今回の研究ではあらかじめ決められた録音機器を用いて複数の音声録音タスクを行う必要があり、研究者が待機できることも必要であるため、研究を実施できる時間と環境が限定されており、また心理検査を含めると少なくとも20分程度は要することからも臨床の間に行うことにやや困難があったことから、対象者を当初予定していたうつ病、統合失調症、健常コントロール各100名ずつという目標値に届かなかった。健常コントロールにはQIDS-SRJが6以上であり軽症のうつ病が疑われる対象が予想より多く、対象が学生や研修医など比較的若年者が多くなりやすかったため、コントロール群の年齢がうつ病群や統合失調症群と比較して若くなりやすかったことや、コントロール群の対象は医療関係者が多いことなど、サンプルの偏りがあることもバイアスとなる。インターネットなどで幅広い対象から多くの音声を収集することで、年齢や性別をマッチングし、今回の結果の再現性を確認する必要がある。このため、今回用いたデバイス以外での収録や、サンプリング周波数などの録音設定の違いがあっても補正することで様々なデバイスに応用できるようにし、今回の結果を元に録音タスクを単純化することが必要である。また、向精神薬による音声への影響についても考慮する必要があるため、今後、服薬歴のない患者の音声を対象に検討する必要がある。

次に、今回はECT前後という短期的な治療効果しかみていないが、一般にうつ病は寛解に至るまでは薬物療法を用いて、ある程度長期的なフォローが必要とされるため、長期的な改善によってMFCCが変化するかどうか、12週、24週、36週などの時点において寛解と判定された時点での結果についても確認する必要がある。また、対象者には維持ECT中の患者もおり、重症度やECTの実施回数のばらつきもあったため、初回ECTに限定するなど対象者の属性を揃え

ることができれば、より変化を捉えやすくなることが考えられる。

統合失調症においては状態像が様々であることや病的体験の影響などから解釈が困難となっており、構造化面接を用いた客観的診断を用いたより正確な抑うつの評価をした上での再評価や、状態像を分けた検討も必要である。また、統合失調症以外の抑うつを呈する疾患との鑑別に利用できるかどうかさらなる検討を要する。

うつ病は病因仮説が様々であるように、現時点では病因が不明であり、今回の対象も DSM-5 による操作的に診断されている。このため、研究 1 における判別分析で、うつ病群と判別されなかったうつ病患者は声道の変化が起こりにくい特性を持つうつ病である可能性や、現時点でうつ病と診断されていても、今後、認知症や双極性障害などの診断に変更となる可能性なども考えられるが、この点は本研究では明らかでない。また、うつ病群と判別された健常コントロールの対象は、何らかの理由でうつ病患者と同様の声道の変化を来している可能性や、今後うつ病を発症する可能性なども考えられるが、この点についても本研究では明らかでない。今回の研究 1 では抑うつ症状の種類よっての関連は十分に検討できていないため、今後 QIDS-SRJ の下位項目との関連を検討したい。

また、使用したマイクロフォンは 800Hz 付近をピークとした中低域の感度が高い周波数特性を持っており (図 3)、これは音声通話を目的としたデバイスであるために、ノイズとなりうる低域や高域が抑えられ、音声が目立ちやすい中域が強調されている可能性が考えられる。3000Hz 付近にも小さなピークがあることも人間の聴覚特性と合致している。このマイクロフォンは汎用品であり計測用のマイクロフォンではないため、その周波数特性からも人間の声への感度が高く、マイクロフォン用のプリアンプリファァーなどの装置を要しないため簡便に用いることができるという利点がある。一方で、周波数特性がフラットでないこと自体がバイアスとなり得ることが欠点と考えられる。このマイクロフォンを使用したことにより、測定用のフラットな周波数特性を持つマイクロフォンと比較し、今回得られた 2000Hz から 3000Hz における影響が強調された可能性が考えられる。このため、マイクの影響との関連や、その補正が可能かどうかについても検討を要する。

MFCC は声道の特性を反映するとされているが、今回得られた MFCC 2 の変化が実際の声道の形態とどのように関連するかは不明である。一方で閉塞性睡眠時無呼吸など夜間に低換気となる状態と音声変化との関連も指摘されているため、終夜ポリソムノグラフィー検査や咽頭・喉頭の MRI と音声との関連の検討も望まれる。

また、今回バイマーカーとなりうる音響特徴量として見出した MFCC 2 はその瞬間の値ではなく発話区間内の平均値として抽出しているため、ある程度の種類の母音が含まれた発話であれば、理論的には読み上げる内容による変化は少ないとは考えられるが、実際に、読み上げる文章の内容によって音響特徴量にどの程度の変化が出るか、どのくらいの長さの音声为好ましい

のかという点も今後評価の必要がある。

今後、長期的な治療による寛解で変化するかどうか、躁うつ病、脳血管性、認知症、パーキンソン病など他の精神疾患に伴う抑うつ状態との鑑別が可能かどうかを検討したい。また、初回の測定時にうつ病を発症していなかった対象が発症した場合の結果を前向き研究で検討することで、発症の予測マーカーとして利用できないかも検討を行いたい。これらの検討によって本研究の成果を深化すれば、音声はうつ病のバイオマーカーとしての地位を確立できると考えられる。そして、精神的な内容に関する電話相談の際に相談内容のみでなく、音声による抑うつや自殺のリスク評価を行うことや、コールセンターなどの顧客対応における抑うつのスクリーニング、スマートフォンや Web ページなどでの精神疾患の鑑別に利用できれば、適切な相談窓口につなぐなどの予防医学的な応用が期待できる。また、他疾患との鑑別に用いることが可能となった場合は、統合失調症の初期、躁うつ病の初回うつ病エピソード、認知症の前駆症状など、抑うつ症状を来すがうつ病との鑑別が難しいケースもある疾患において、鑑別診断の補助として用いることも期待される。

## 結論

数字読み上げ音声において、声の聞き取りやすさに関連すると考えられるメル周波数ケプストラム係数 (MFCC) 2 が抑うつ状態に変化し、これがうつ病を統合失調症などと鑑別できる診断バイオマーカーとなる可能性が示唆された。今後、より多くのサンプルサイズで、長期のフォローアップや状態像などのサブタイプでの差、実際の声道の変化、他の疾患での変化に関するさらなる検討が必要である。

## 謝辞

今回の研究活動においてご指導いただきました、筑波大学医学医療系精神医学分野の新井哲明教授、太刀川弘和准教授、根本清貴准教授、静岡大学大学院情報学研究科の西村雅史教授、IBM Research – Tokyo の立花隆輝様、長野徹様、鈴木雅之様に深謝いたします。

データ収集にご協力いただいた、茨城県立こころの医療センターの土井永史院長、山田典子様、有朋会栗田病院の栗田裕文理事長、安部秀三院長、医療法人イプシロンの高尾哲也理事長に深謝いたします。

臨床を続けながら研究をさせていただいた、有朋会栗田病院医局の皆様、茨城県立こころの医療センター医局の皆様、また研究を支えてくださった研究室の皆様、親愛なる家族にも深謝いたします。

## 文献

- Abbott, C.C., Jones, T., Lemke, N.T., Gallegos, P., McClintock, S.M., Mayer, A.R., Bustillo, J., Calhoun, V.D., 2014. Hippocampal structural and functional changes associated with electroconvulsive therapy response. *Transl. Psychiatry*.
- Akkaralaertsest, T., Yingthawornsuk, T., 2015. Comparative Analysis of Vocal Characteristics in Speakers with Depression and High-Risk Suicide. *Int. J. Comput. Theory Eng.* 7, 448–452.
- Alpert, M., Pouget, E.R., Silvia, R.R., 2001. Reflections of depression in acoustic measures of the patient's speech. *J. Affect. Disord.* 66, 59–69.
- American Psychiatric Association., American Psychiatric Association. DSM-5 Task Force., 2013. Diagnostic and statistical manual of mental disorders : DSM-5. American Psychiatric Association.
- Atkinson, A.J., Colburn, W.A., DeGruttola, V.G., DeMets, D.L., Downing, G.J., Hoth, D.F., Oates, J.A., Peck, C.C., Schooley, R.T., Spilker, B.A., Woodcock, J., Zeger, S.L., 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69, 89–95.
- Bartels, S.J., Drake, R.E., 1988. Depressive symptoms in schizophrenia: comprehensive differential diagnosis. *Compr. Psychiatry* 29, 467–83.
- Bele, I.V., 2006. The Speaker's Formant. *J. Voice* 20, 555–578.
- Belvederi Murri, M., Amore, M., Calcagno, P., Respino, M., Marozzi, V., Masotti, M., Bugliani, M., Innamorati, M., Pompili, M., Galderisi, S., Maj, M., 2016. The "Insight Paradox" in Schizophrenia: Magnitude, Moderators and Mediators of the Association Between Insight and Depression. *Schizophr. Bull.* 42, 1225–1233.
- Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., Snyder, P.J., 2004. Voice acoustical measurement of the severity of major depression. *Brain Cogn.* 56, 30–5.
- Coplan, J.D., Gopinath, S., Abdallah, C.G., Berry, B.R., 2014. A Neurobiological Hypothesis of Treatment-Resistant Depression " Mechanisms for Selective Serotonin Reuptake Inhibitor Non-Efficacy. *Front. Behav. Neurosci.* 8, 1–16.
- Cuijpers, P., Van Straten, A., Smits, N., Smit, F., 2006. Screening and early psychological intervention for depression in schools: Systematic review and meta-analysis. *Eur. Child Adolesc. Psychiatry* 15, 300–307.
- Cullen, K., Klimes-Dougan, B., Kumra, S., Schulz, S.C., 2009. Paediatric major depressive disorder: Neurobiology and implications for early intervention. *Early Interv. Psychiatry* 3,



178–188.

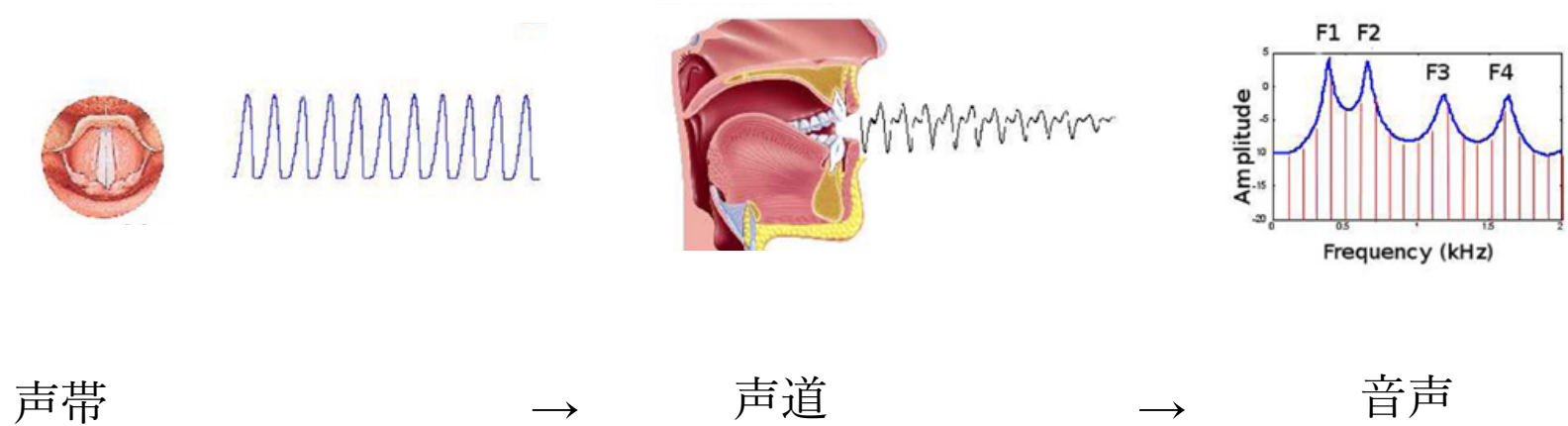
- Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F., 2015a. A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* 71, 10–49.
- Cummins, N., Sethu, V., Epps, J., Schnieder, S., Krajewski, J., 2015b. Analysis of acoustic space variability in speech affected by depression. *Speech Commun.* 75, 27–49.
- Davis, S.B., Mermelstein, P., 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Signal Process.* 28, 357–366.
- de Voogd, J.N., Wempe, J.B., Koëter, G.H., Postema, K., van Sonderen, E., Ranchor, A. V., Coyne, J.C., Sanderman, R., 2009. Depressive Symptoms as Predictors of Mortality in Patients With COPD. *Chest* 135, 619–625.
- Depression in adults: recognition and management | Guidance and guidelines | NICE [https://www.nice.org.uk/guidance/CG90], 2009.
- Dunn, E.C., Brown, R.C., Dai, Y., Rosand, J., Nugent, N.R., Amstadter, A.B., Smoller, J.W., 2015. Genetic determinants of depression: Recent findings and future directions. *Harv Rev Psychiatry* 23, 1–18.
- Eyben, F., Woellmer, M., Schuller, B., 2010. openSMILE the Munich open Speech and Music Interpretation by Large Space Extraction toolkit.
- Fischer, S., Strawbridge, R., Vives, A.H., Cleare, A.J., 2017. Cortisol as a predictor of psychological therapy response in depressive disorders: Systematic review and meta-analysis. *Br. J. Psychiatry* 210, 105–109.
- Flint, A.J., Black, S.E., Taylor, I.C., Galley, G.F., Levinton, C., 1992. Acoustic analysis in the differentiation of Parkinson's disease and major depression. *J. Psycholinguist. Res.* 21, 383–399.
- Fu, C.H.Y., Steiner, H., Costafreda, S.G., 2013. Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiol. Dis.* 52, 75–83.
- Fujisawa, D., 2010. Assessment scales of cognitive behavioral therapy. *Japanese J. Clin. psychiatry* 39, 839–850.
- Haapakoski, R., Mathieu, J., Ebmeier, K.P., Alenius, H., Kivimäki, M., 2015. Cumulative meta-analysis of interleukins 6 and  $1\beta$ , tumour necrosis factor  $\alpha$  and C-reactive protein in patients with major depressive disorder. *Brain. Behav. Immun.* 49, 206–15.

- Husain, M.M., Rush, A.J., Fink, M., Knapp, R., Petrides, G., Rummans, T., Biggs, M.M., O'Connor, K., Rasmussen, K., Litle, M., Zhao, W., Bernstein, H.J., Smith, G., Mueller, M., McClintock, S.M., Bailine, S.H., Kellner, C.H., 2004. Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): A consortium for research in ECT (CORE) report. *J. Clin. Psychiatry* 65, 485–491.
- Inada, T., Inagaki, A., 2015. Psychotropic dose equivalence in Japan. *Psychiatry Clin. Neurosci.* 69, 440–447.
- Janakiramaiah, N., Gangadhar, B.N., Naga Venkatesha Murthy, P.J., Harish, M.G., Subbakrishna, D.K., Vedamurthachar, A., 2000. Antidepressant efficacy of Sudarshan Kriya Yoga (SKY) in melancholia: A randomized comparison with electroconvulsive therapy (ECT) and imipramine. *J. Affect. Disord.* 57, 255–259.
- Joshi, S.H., Espinoza, R.T., Pirnia, T., Shi, J., Wang, Y., Ayers, B., Leaver, A., Woods, R.P., Narr, K.L., 2016. Structural plasticity of the hippocampus and amygdala induced by electroconvulsive therapy in major depression. *Biol. Psychiatry* 79, 282–292.
- Kaufman, J., DeLorenzo, C., Choudhury, S., Parsey, R. V., 2016. The 5-HT<sub>1A</sub> receptor in Major Depressive Disorder. *Eur Neuropsychopharmacol.* 26, 397–410.
- Ladd, D.R., 1980. *Structure of Intonational Meaning: Evidence from English*. Indiana University Press.
- Leino, T., Laukkanen, A.M., Radolf, V., 2011. Formation of the actor's/speaker's formant: A study applying spectrum analysis and computer modeling. *J. Voice* 25, 150–158.
- Masuda, K., Nakanishi, M., Okamoto, K., Kawashima, C., Oshita, H., Inoue, A., Takita, F., Izumi, T., Ishitobi, Y., Higuma, H., Kanehisa, M., Ninomiya, T., Tanaka, Y., Akiyoshi, J., 2017. Different functioning of prefrontal cortex predicts treatment response after a selective serotonin reuptake inhibitor treatment in patients with major depression. *J. Affect. Disord.* 214, 44–52.
- Maurer, J., Rebbapragada, V., Borson, S., Goldstein, R., Kunik, M.E., Yohannes, A.M., Hanania, N.A., 2008. Anxiety and Depression in COPD. *Chest* 134, 43S–56S.
- Ministry of Health Labour and Welfare (Japan), 2014. Patient Survey 2014 1–33.
- Mitrović, D., Zeppelzauer, M., Breiteneder, C., 2010. Features for Content-Based Audio Retrieval. *Adv. Comput.* 78, 71–150.
- Molendijk, M.L., Spinhoven, P., Polak, M., Bus, B.A.A., Penninx, B.W.J.H., Elzinga, B.M., 2014. Serum BDNF concentrations as peripheral manifestations of depression: evidence from a systematic review and meta-analyses on 179 associations (N=9484). *Mol. Psychiatry* 19,

- 791–800.
- Mundt, J.C., Vogel, A.P., Feltner, D.E., Lenderking, W.R., 2012. Vocal acoustic biomarkers of depression severity and treatment response. *Biol. Psychiatry* 72, 580–7.
- Nemesysco, 2007. Nemesysco's Voice Analysis Technology Helps Detect Mental Stress. [<http://nemesysco.com/press-releases-detect-mental-stress>].
- Nilsonne, A., Sundberg, J., Ternström, S., Askenfelt, A., 1988. Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *J. Acoust. Soc. Am.* 83, 716–728.
- Rush, A.J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., Markowitz, J.C., Ninan, P.T., Kornstein, S., Manber, R., Thase, M.E., Kocsis, J.H., Keller, M.B., 2003. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Depression* 54, 573–583.
- Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Commun.* 54, 543–565.
- Schuller, B., Steidl, S., Batliner, A., 2009. The INTERSPEECH 2009 emotion challenge. *INTERSPEECH-2009* 312–315.
- Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Christian, M., Narayanan, S., 2010. The INTERSPEECH 2010 paralinguistic challenge. *INTERSPEECH-2010* 2794–2797.
- Schwartz, D.J., Karatinos, G., 2007. For individuals with obstructive sleep apnea, institution of CPAP therapy is associated with an amelioration of symptoms of depression which is sustained long term. *J. Clin. Sleep Med.* 3, 631–635.
- Solomon, C., Valstar, M.F., Morriss, R.K., Crowe, J., 2015. Objective Methods for Reliable Detection of Concealed Depression. *Front. ICT* 2, 1–16.
- Strawbridge, R., Young, A.H., Cleare, A.J., 2017. Biomarkers for depression: Recent insights, current challenges and future prospects. *Neuropsychiatr. Dis. Treat.* 13, 1245–1262.
- Streeter, L.A., Macdonald, N.H., Apple, W., Krauss, R.M., Galotti, K.N., 1983. Acoustic and perceptual indicators of emotional stress. *J. Acoust. Soc. Am.* 73, 1354–1360.
- Sundberg, J., 2001. Level and center frequency of the singer's formant. *J. Voice* 15, 176–186.
- Teintipsiri, T., Yingthawornsuk, T., 2014. Comparative Study of MFCC in Categorized Speech Samples Related Depression. *Int. Conf. Intell. Syst. Data Min. Inf. Technol.*

- (ICIDIT'2014), April 21-22, 2014 Bangkok.
- Tendolkar, I., van Beek, M., van Oostrom, I., Mulder, M., Janzing, J., Voshaar, R.O., van Eijndhoven, P., 2013. Electroconvulsive therapy increases hippocampal and amygdala volume in therapy refractory depression: A longitudinal pilot study. *Psychiatry Res. - Neuroimaging* 214, 197–203.
- Tolkmitt, F.J., Scherer, K.R., 1986. Effect of Experimentally Induced Stress on Vocal Parameters. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 302–313.
- WHO, 2017a. Depression [<http://www.who.int/mediacentre/factsheets/fs369/en/>].
- WHO, 2017b. Global Health Estimates [[http://who.int/healthinfo/global\\_burden\\_disease/%0D%0Aestimates/en/%0D%0A](http://who.int/healthinfo/global_burden_disease/%0D%0Aestimates/en/%0D%0A)].
- Wise, T., Cleare, A.J., Herane, A., Young, A.H., Arnone, D., 2014. Diagnostic and therapeutic utility of neuroimaging in depression: An overview. *Neuropsychiatr. Dis. Treat.* 10, 1509–1522.
- Yinghua Zhu, Yoon-Chul Kim, Proctor, M.I., Narayanan, S.S., Nayak, K.S., 2013. Dynamic 3-D Visualization of Vocal Tract Shaping During Speech. *IEEE Trans. Med. Imaging* 32, 838–848.
- 株式会社日立システムズ, 2017. 音声こころ分析サービス [<http://www.hitachi-systems.com/sp/mimosys/>].
- 厚生労働省, 2015. ストレスチェック等の職場におけるメンタルヘルス対策・過重労働対策等 [<http://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/>].
- 田口高也, 根本清貴, 太刀川弘和, 長野徹, 立花隆輝, 西村雅史, 新井哲明, 朝田隆, 2014. OpenEARを用いた音声による心理的ストレス検出の試み. *精神医学* 56, 1027–1034.

## 図表



(ブザーのような音)

(響きを付加)

図 : Moulin-Frier , et al. 2014 改変

name	energy	mfcc 1	mfcc 2	mfcc 3
subject01	1.66E-02	2.96E-04	1.63E-02	2.86E+02
subject02	9.03E-03	7.05E-04	8.33E-03	1.70E+01
subject03	7.51E-03	1.17E-03	6.34E-03	2.06E+02
subject04	4.38E-02	1.01E-03	4.28E-02	1.77E+02
subject05	1.27E-02	5.85E-04	1.21E-02	4.40E+01

↙ OpenSMILE

(音響特徴量を抽出)

(抽出されたデータの例)

図 1. 音声の成り立ちと音響特徴量の抽出



図 2. 使用したタブレット端末と受話器型ヘッドセット

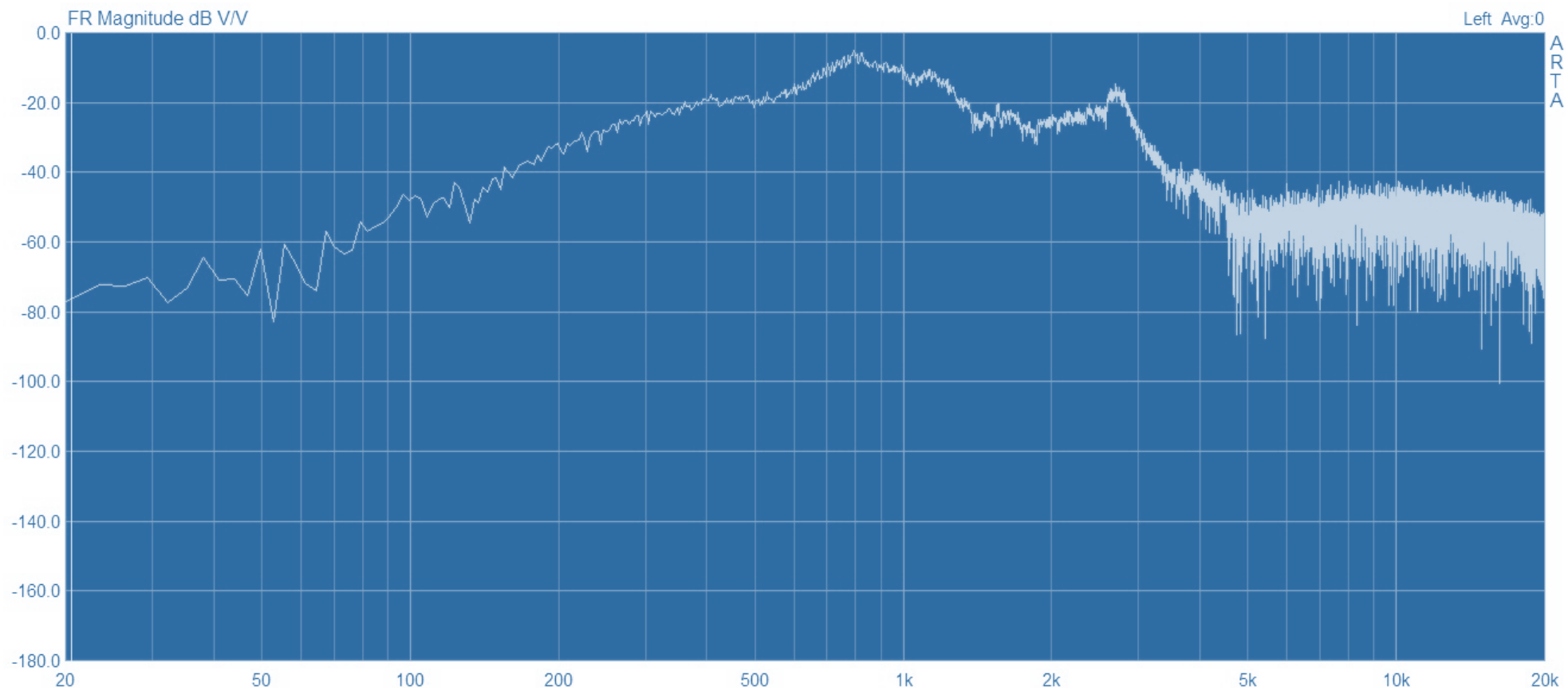
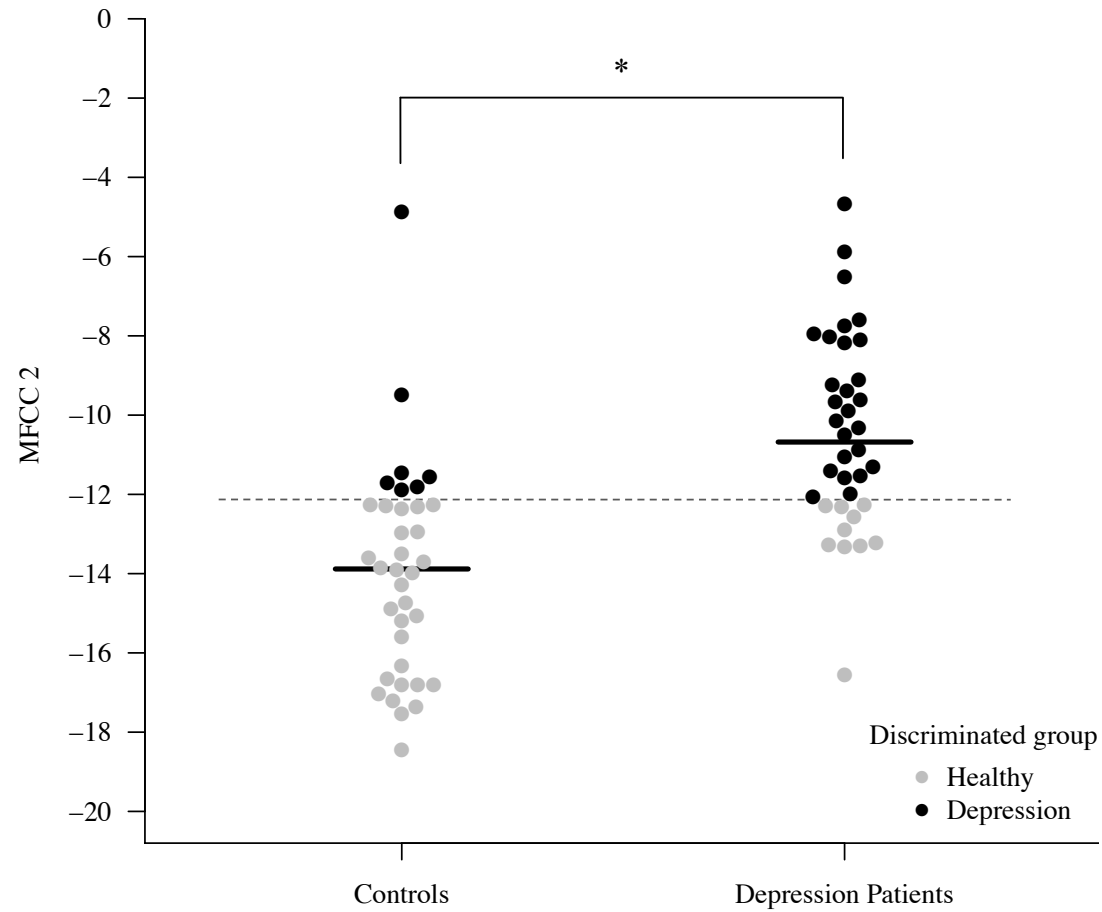


図 3. 使用したマイクロフォンの周波数特性





\* p < 0.001

図 4. うつ病群とコントロール群における MFCC 2

黒の点はうつ病群と判別された対象者であり、灰色の点はコントロール群と判別された対象者である。各群の横線は平均値を示す。

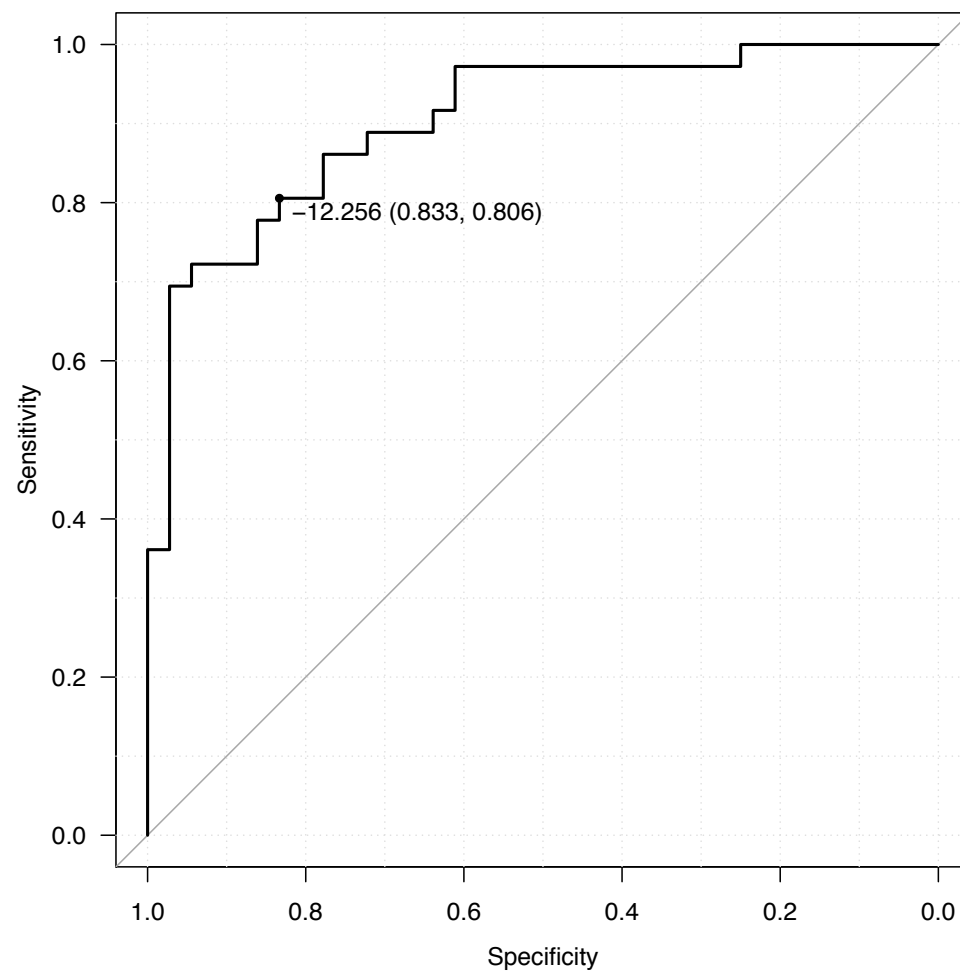


図 5. MFCC 2 の Receiver Operatorating Characteristic (ROC)曲線

Area under curve (AUC) は 0.88 であった。

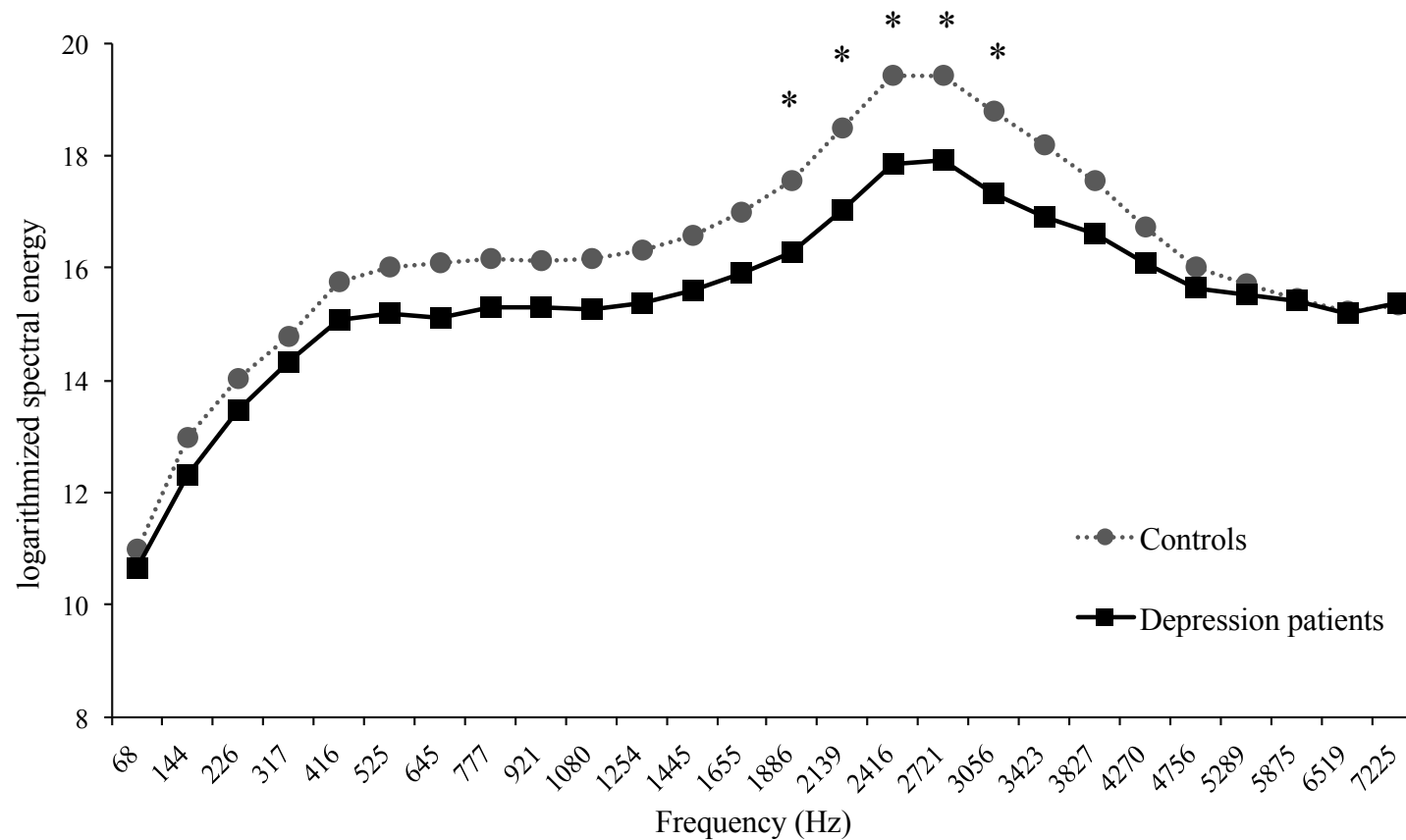


図6. うつ病群とコントロール群の音声スペクトルの比較

水平軸はメル尺度（人間が感じる音程の聴覚特性に合わせた周波数尺度）に基づいた周波数スケールを示し、垂直軸は対数エネルギーを示し、値は各群の平均値を示す。

うつ病群の音声は2000 Hzから3000 Hzにおいてコントロール群よりエネルギーが低下していた。

\*  $p < 0.001$  (Bonferroniの補正後)

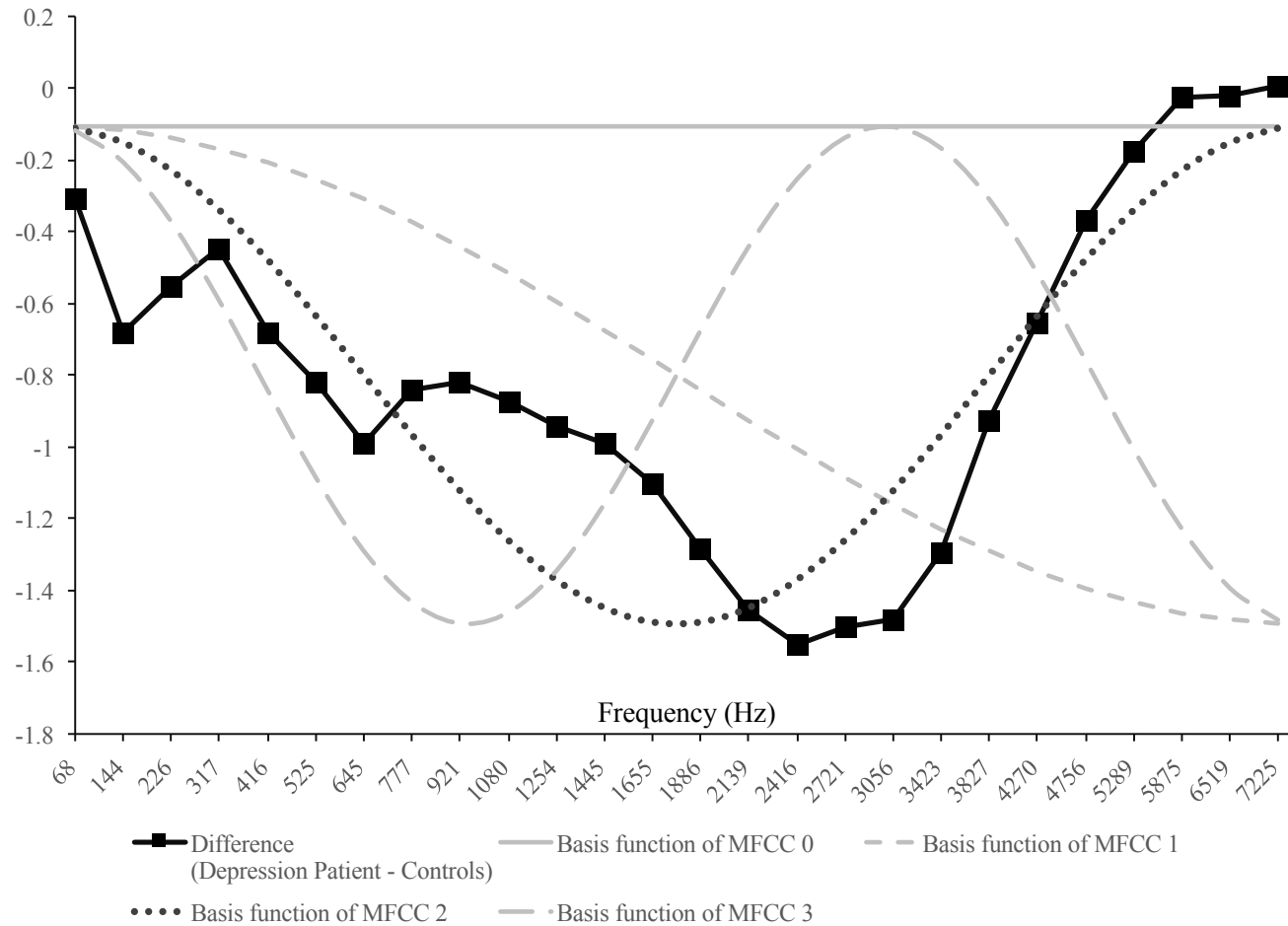


図 7. MFCC 2 のうつ病群とコントロール群における差分、および MFCC 0 から 3 の離散コサイン変換における基底  
 水平軸はメル尺度に基づいた周波数スケール、垂直軸はうつ病群とコントロール群の MFCC 2 の差分（対数エネルギーの差分）を示す。  
 MFCC 2 の差分は MFCC 2 の基底と類似しているため、うつ病群とコントロール群において MFCC 2 に有意な差を認めたと考えられる。

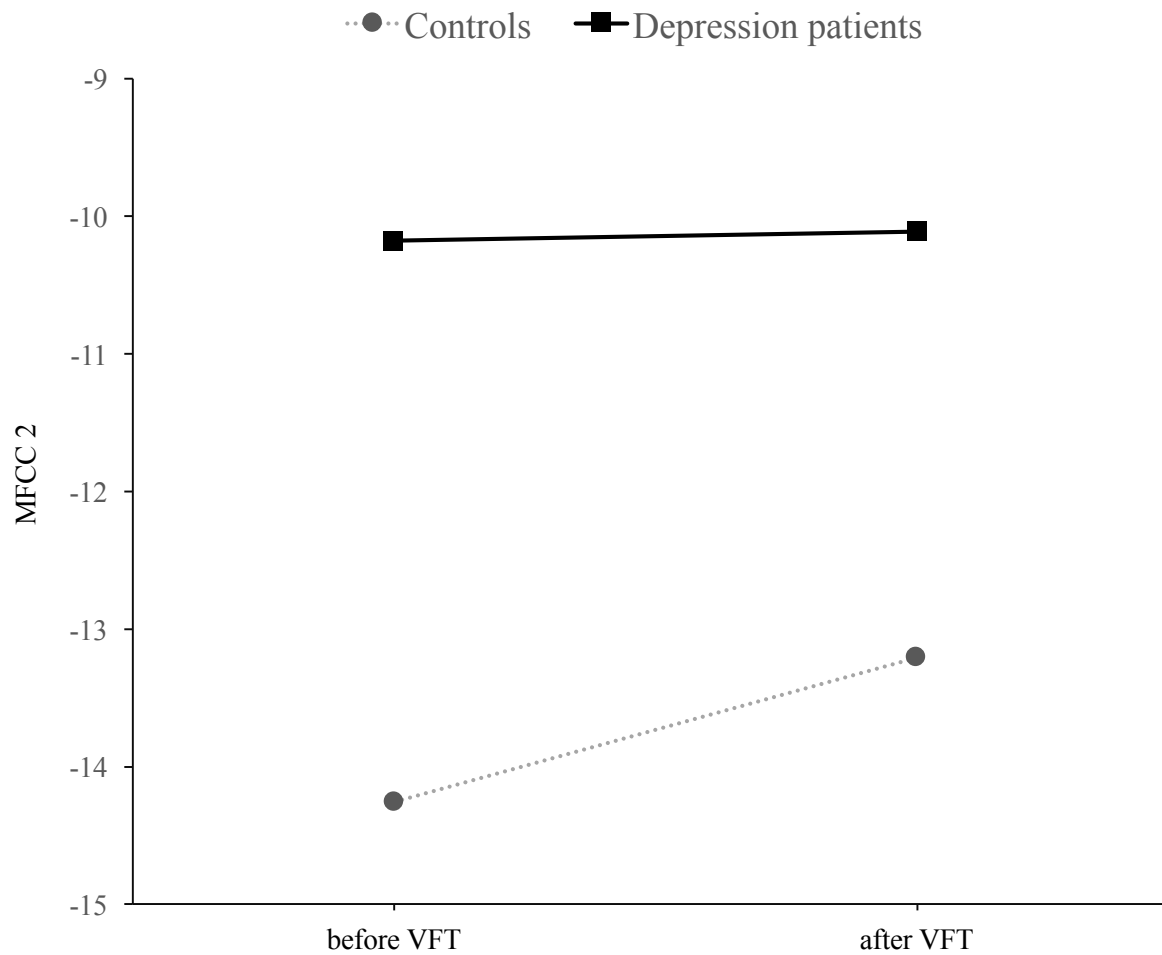


図 8. 語流暢性課題の前後における MFCC 2

語流暢性課題前後における MFCC 2 の交互作用は認めなかった。

VFT: Verbal fluency test (語流暢性課題)

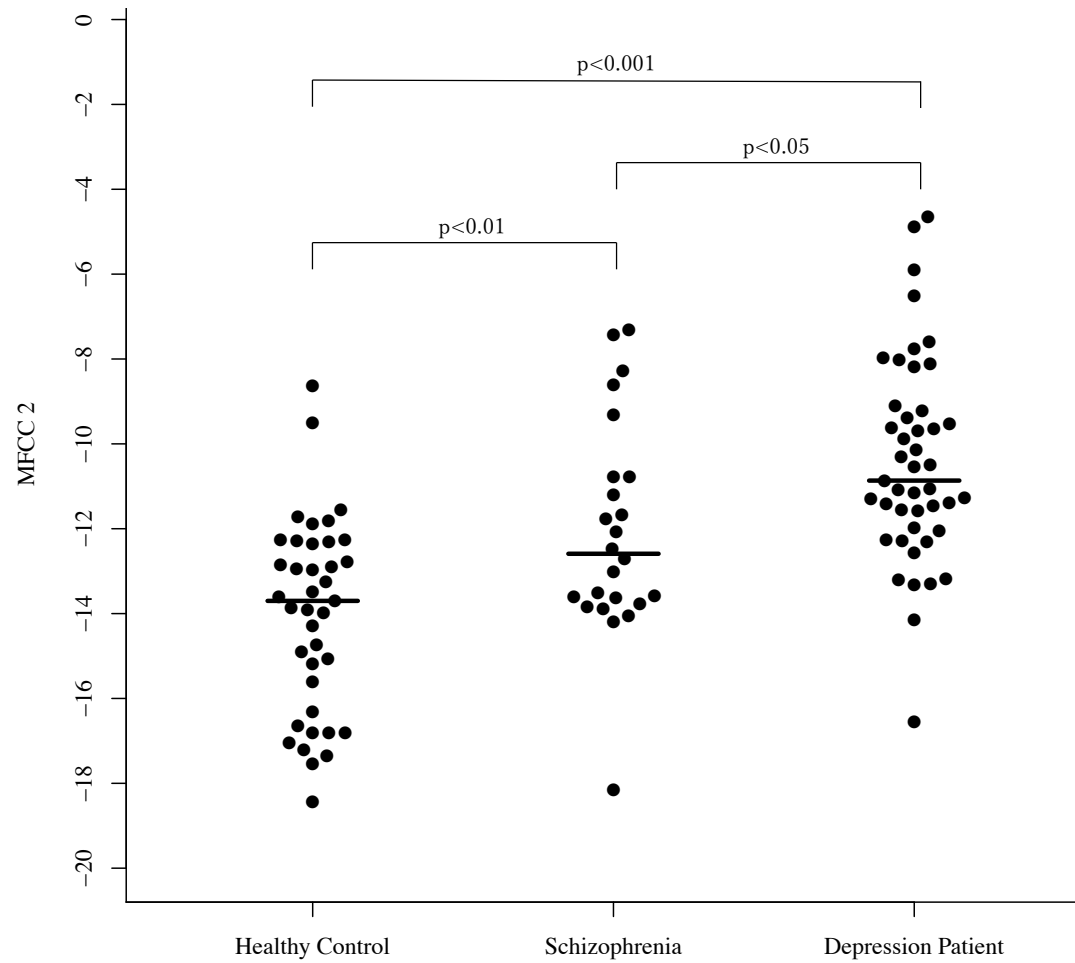


図 9. コントロール群、統合失調症群、うつ病群の比較

各群の横線は平均値を示す。

統合失調症群はうつ病群とコントロール群の中間を示した。

表 1. 音響特徴量の例

音響特徴量	説明	意義、応用
基本周波数	最も低い周波数ピーク	音程
フォルマント	基本周波数より高い複数の周波数のピーク	音色や音質が影響する声道の共振周波数
音量の二乗平均平方根	時間で変化する音量の評価方法	音の大きさ
高調波対ノイズ比	音声の付加ノイズ量の評価方法	音声の濁り
ジッター	時間軸方向での揺らぎ	音程変化
シマー	振幅方向での揺らぎ	音量変化
ゼロクロス率	音声波が時間軸を通過する頻度	音声の明瞭さ
線形予測符号 (LPC)	予測したフォルマントを除去し、残った部分の強さと周波数を予測	音声圧縮、音声認識
知覚的線形予測 (PLP)	LPC を人間の聴覚特性に合わせたもの	音声圧縮、音声認識に応用
メル周波数ケプストラム係数 (MFCC)	音声スペクトル成分を概形化	音声認識に応用

表 2. 対象

	Depression patients (mean ± SD)	Controls (mean ± SD)	<i>p</i>
<i>n</i> (male/female)	36 (22/14)	36 (16/20)	n.s.
Age (years)	44.0 ± 16.3	38.0 ± 10.4	n.s.
QIDS-SRJ	11.7 ± 6.2	2.5 ± 1.8	<0.001
Disease duration (years)	7.00 ± 5.73		
Antidepressants (mg)	114.0 ± 72.4		
Antipsychotics (mg)	54.1 ± 117.5		
Anxiolytics (mg)	9.9 ± 11.0		

SD: standard deviation.

QIDS-SRJ: Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version.

n.s.: not significant.

向精神薬は抗精神病薬、抗うつ薬、抗不安薬（睡眠薬を含む）に分類し、それぞれ等価換値を示した（抗精神病薬：クロルプロマジン換算、抗うつ薬：イミプラミン換算、抗不安薬：ジアゼパム換算）。（稲田、稲垣、2015）。



表 3. うつ病群とコントロール群の比較

Acoustic features	Depression patients (SD)		Controls (SD)		<i>p</i> -value	
	<i>n</i> = 36	Male ( <i>n</i> = 22) Female ( <i>n</i> = 14)	<i>n</i> = 36	Male ( <i>n</i> = 16) Female ( <i>n</i> = 20)	Total	Male Female
Root mean square energy	5.23×10 <sup>-3</sup> (3.49×10 <sup>-3</sup> )	5.63×10 <sup>-3</sup> (4.01×10 <sup>-3</sup> ) 4.59×10 <sup>-3</sup> (2.49×10 <sup>-3</sup> )	7.31×10 <sup>-3</sup> (7.70×10 <sup>-3</sup> )	9.36×10 <sup>-3</sup> (1.06×10 <sup>-2</sup> ) 5.67×10 <sup>-3</sup> (3.67×10 <sup>-3</sup> )	n.s.	n.s. n.s.
Zero crossing rate	1.43×10 <sup>-3</sup> (1.97×10 <sup>-2</sup> )	1.48×10 <sup>-3</sup> (2.13×10 <sup>-2</sup> ) 1.35×10 <sup>-3</sup> (1.44×10 <sup>-2</sup> )	1.49×10 <sup>-3</sup> (1.74×10 <sup>-2</sup> )	1.55×10 <sup>-3</sup> (1.78×10 <sup>-2</sup> ) 1.45×10 <sup>-3</sup> (1.63×10 <sup>-2</sup> )	n.s.	n.s. n.s.
Harmonics to noise ratio	0.351 (0.0589)	0.323 (0.0213) 0.396 (0.0673)	0.370 (0.0536)	0.341 (0.0448) 0.393 (0.0495)	n.s.	n.s. n.s.
Fundamental frequency	32.54 (31.1)	14.1 (9.62) 61.4 (31.4)	35.5 (27.2)	18.5 (16.4) 49.6 (26.3)	n.s.	n.s. n.s.
MFCC 1	-10.3 (1.65)	-10.3 (1.92) -10.2 (1.16)	-10.2 (1.94)	-9.73 (2.04) -10.5 (1.83)	n.s.	n.s. n.s.
MFCC 2	-10.2 (2.56)	-10.3 (2.43) -10.1 (2.84)	-14.3 (2.16)	-14.0 (2.18) -14.5 (2.84)	<0.001	<0.001 <0.001
MFCC 3	-0.04 (3.62)	-0.682 (3.38) -1.18 (3.80)	2.65 (1.99)	3.45 (1.76) 2.01 (1.96)	n.s.	n.s. n.s.
MFCC 4	-7.84 (3.76)	-6.16 (3.26) -10.5 (2.94)	-8.89 (3.23)	-7.78 (3.01) -9.78 (3.19)	n.s.	n.s. n.s.
MFCC 5	-11.9 (2.22)	-11.7 (1.96) -12.2 (2.63)	-13.6 (3.60)	-13.8 (3.99) -13.5 (2.63)	n.s.	n.s. n.s.
MFCC 6	-2.48 (3.05)	-1.59 (3.17) -3.90 (2.29)	-1.24 (3.25)	-0.145 (3.47) -2.11 (2.84)	n.s.	n.s. n.s.
MFCC 7	-3.64 (3.80)	-1.95 (3.37) -6.29 (2.85)	-3.41 (2.39)	-1.96 (2.23) -4.58 (1.83)	n.s.	n.s. n.s.
MFCC 8	-5.80 (3.27)	-5.84 (3.22) -5.74 (3.46)	-5.06 (2.69)	-4.91 (2.71) -5.18 (2.74)	n.s.	n.s. n.s.
MFCC 9	-2.91 (3.22)	-2.16 (3.27) -4.09 (2.86)	-2.62 (2.79)	-2.01 (3.05) -3.11 (2.53)	n.s.	n.s. n.s.
MFCC 10	-0.532 (2.67)	0.658 (2.21) -2.40 (2.28)	-0.408 (2.44)	1.28 (2.06) -1.76 (1.83)	n.s.	n.s. n.s.
MFCC 11	-2.33 (2.80)	-0.706 (2.05) -4.89 (1.67)	-2.57 (2.29)	-1.77 (2.46) -3.21 (1.99)	n.s.	n.s. n.s.
MFCC 12	-2.06 (1.97)	-1.96 (1.97) -2.21 (2.02)	-2.80 (1.80)	-3.01 (1.75) -2.62 (1.87)	n.s.	n.s. n.s.

SD: standard deviation

n.s.: not significant after Bonferroni correction

表 4. MFCC 2 における判別分析の結果

Groups	Results	
	Depression patients (male/female)	Controls (male/female)
Depression patients ( $n = 36$ )	28 (18/10)	8 (4/4)
Controls ( $n = 36$ )	5 (3/2)	31 (13/18)

Stepwise discriminant analysis with MFCCs revealed that only MFCC 2 contributed to the discrimination (Wilks'  $l = 0.567$ ;  $c^2 = 39.456$ ,  $df = 1$ ,  $p < 0.001$ ).

表 5. MFCC 2 の重回帰分析

Variable	<i>r</i>			$\beta$	<i>t</i>	<i>p-value</i>
	MFCC 2	Gender	Age			
Group	0.658	-0.167	0.215	0.637	6.673	$5.48 \times 10^{-9}$
Age	0.197	0.209		0.068	0.708	n.s.
Gender	-0.131			-0.039	-0.404	n.s.

n.s.: not significant

表 6. 対象

	Depression patients (mean ± SD)	<i>p</i>
<i>n</i> (male/female)	10 (5/5)	n.s.
Age (years)	55.0 ± 19.0	n.s.
Disease duration (years)	4.9 ± 4.8	
Antidepressants (mg)	134.5 ± 188.8	
Antipsychotics (mg)	123.0 ± 120.0	
Anxiolytics (mg)	5.12 ± 4.65	

SD: standard deviation

QIDS-SRJ: Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version

n.s.: not significant

向精神薬は抗精神病薬、抗うつ薬、抗不安薬（睡眠薬を含む）に分類し、それぞれ等価換値を示した（抗精神病薬：クロルプロマジン換算、抗うつ薬：イミプラミン換算、抗不安薬：ジアゼパム換算）。

表 7. ECT 前後の QIDS-J

ECT 後 - ECT 前	差の平均	差の標準偏差	<i>t</i>
QIDS-J 総スコア	6.80*	5.79	3.71
QIDS-J 睡眠	-0.20	0.42	-1.50
QIDS-J 抑うつ	0.70	1.16	1.91
QIDS-J 食欲体重	0.40	1.71	0.74
QIDS-J 集中力決断	1.30*	1.06	3.88
QIDS-J 自己評価	1.10	1.10	3.16
QIDS-J 希死念慮	0.50	0.85	1.86
QIDS-J 興味	1.10	1.29	2.70
QIDS-J 疲労	1.30	1.16	3.55
QIDS-J 精神運動状態	0.60	0.70	2.71

\*:  $p < 0.05$  after Bonferroni correction

QIDS-J: Quick Inventory of Depressive Symptomatology, Japanese version

表 8. 対象

	Schizophrenia patients (mean ± SD)	Depression patients (mean ± SD)	Controls (mean ± SD)	<i>p</i>
<i>n</i> (male/female)	24 (9/15)	45 (27/18)	39 (16/23)	n.s. †
Age (years)	40.3 ± 12.4	43.4 ± 16.9	37.5 ± 11.7	n.s. ††
QIDS-SRJ	11.4 ± 5.4	11.3 ± 5.9	2.89 ± 2.2	
Disease duration (years)	14.3 ± 13.0	8.4 ± 7.5		
Antidepressants (mg)	20.1 ± 68.5	122.0 ± 87.7		
Antipsychotics (mg)	669.1 ± 493.9	38.5 ± 107.0		
Anxiolytics (mg)	11.1 ± 13.9	10.0 ± 10.0		

SD: standard deviation

QIDS-SRJ: Quick Inventory of Depressive Symptomatology – Self-Report, Japanese version

† :  $\chi^2$ 検定

†† : 一元配置分散分析

n.s.: not significant

## 参考論文