CK-007

発話音声に基づく FTLD・ALS の簡易検出における音響・言語モダリティ混合の検討

Proposal of Acoustic-Linguistic Modality Fusion in Easy Screening of FTLD/ALS Based on Speech

伊藤 有生[1)] 加藤 昇平[2)] 坂口 功一[2)] 佐久間 拓人[2)] 大嶽 れい子[3)]
Yuki Ito Shohei Kato Koichi Sakaguchi Takuto Sakuma Reiko Ohdake

桝田 道人[4)] 渡辺 宏久[3)]
Michihito Masuda Hirohisa Watanabe

1 はじめに

1.1 研究背景

現在，日本は超高齢社会を迎え認知症患者は増加傾向にあり，2025 年には約 700 万人の人が認知症を罹患すると予想されている [1]．こうした中，厚生労働省は認知症高齢者等に優しい社会の実現を目指し「認知症施策推進総合戦略（新オレンジプラン）」を策定した．この戦略の柱の 1 つに「認知症の容態に応じた適時・適切な医療・介護等の提供」を掲げており，認知症の早期診断・対応のための体制整備が急がれている．

認知症には様々な基礎疾患が存在し [2]，それぞれ病態や症状が異なるため，適切に治療するには基礎疾患の正確な診断が求められる．しかし，認知症の基礎疾患の中には非専門医による臨床診断が困難なものが存在する．前頭側頭葉変性症（Frontotemporal Lobar Degeneration: FTLD）[3] はそのうちの 1 つである．FTLD は行動障害や言語障害 などが緩徐に進行する神経変性疾患である．FTLD は指定難病に登録されており，他の認知症に比べて症例数が少なく非専門医による臨床診断が困難である．そのため非専門医の FTLD の診断を支援する簡易スクリーニングツールが求められている．

1.2 関連研究

認知症の臨床診断には血液検査，脳画像検査などがあるが，これらの検査は費用が高額であり侵襲性が高いため，患者の負担が大きい．一方で，認知症の簡易検査には，MMSE（Mini-Mental State Examination）[4] や，HDS-R（改定長谷川式簡易知能評価スケール）[5] などの神経心理学検査があり，主に医療現場において用いられている．しかしながら，簡易検査であっても検査に 5〜20 分程度の時間を要するため，医師の時間的負担が大きい．こうした理由から，FTLD の診断支援を目的とするスクリーニングツールは非侵襲かつ簡易的に実施できる方法が望ましい．

1） 名古屋工業大学 工学部 情報工学科
Dept. of Computer Science, Nagoya Institute of Technology
2） 名古屋工業大学 大学院工学研究科 工学専攻 情報工学系プログラム
Dept. of Engineering, Graduate School of Engineering, Nagoya Institute of Technology
3） 藤田医科大学 医学部 脳神経内科学
Department of Neurology, Fujita Medical University School of Medicine
4） 名古屋大学 大学院医学系研究科 神経内科
Department of Neurology, Fujita Medical University School of Medicine

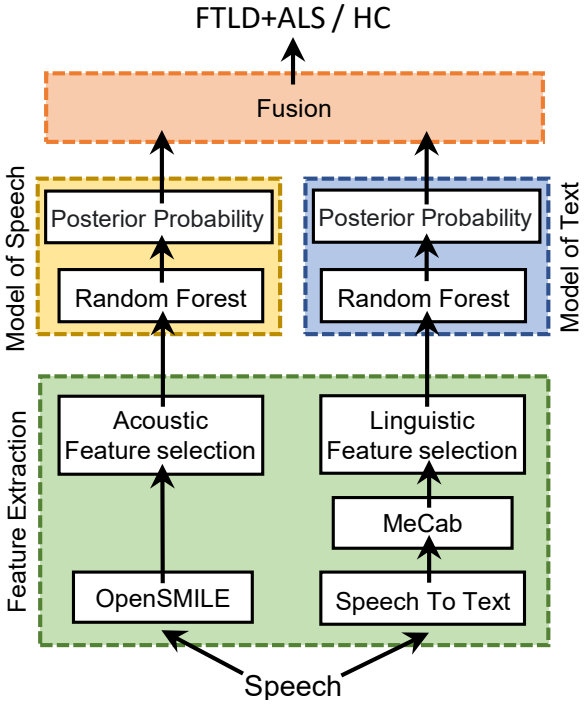

図 1: 提案モデルの概観

近年では，非侵襲かつ簡易的な検査方法として高齢者の音声発話に着目したスクリーニングが研究されている．Kato ら [6] は，日時見当識の応答発話音声から抽出した音韻特徴を用いて健常者と軽度のアルツハイマー病（mAD），軽度認知機能障害を鑑別する手法を提案し，mAD 検出における日時見当識の有効性を確認した．また花井ら [7] は FTLD の言語障害に着目し，自発話課題の回答音声から音響特徴と言語特徴を抽出して健常者・アルツハイマー病・FTLD の鑑別を検証し，言語特徴が FTLD 検出に有効だと示唆している．一方で，日時見当識や自発話課題は認知負荷が高く，発話をためらう場合がある．そこで，坂口ら [8] は認知負荷が低く，発話を誘発しやすい 16 種の音読課題からそれぞれ音響特徴と言語特徴を抽出し，FTLD・ALS を検出する鑑別モデルを提案した．しかしながら，坂口らの提案手法は，音響特徴と言語特徴を結合して学習しており，モダリティの違いを考慮していない．そこで本稿ではモダリティごとに学習器を生成し，予測値を融合することで，音響と言語のモダリティの違いを考慮した FTLD・ALS の鑑別を試みる．

表 1: 実験協力者

	男性	女性	合計人数	平均年齢	MMSE(/30)	WAB(/100)
HC	28	53	81	68.9 ± 8.1	28.8 ± 1.1	96.8 ± 2.4
FTLD+ALS	48	39	87	68.9 ± 9.8	24.1 ± 5.4	88.2 ± 12.9

表 2: 音響特徴量

特徴量	説明
RMS energy	エネルギーの二乗平均平方根値
MFCC	メル周波数ケプストラム係数（12 次元）
Pcm zcr	ゼロ交差率
Voice prob	音が声である確率
F0	基本周波数
* de	特徴量 * の一次微分値

表 3: 統計量

素性値名	説明
max	データの最大値
min	データの最小値
range	最大値と最小値の差
maxPos	最大値を出力した位置
minPos	最小値を出力した位置
amean	算術平均
linregc1	線形近似の勾配度
linregc2	線形近似のオフセット
linregcerrQ	線形近似の二乗誤差
stddev	標準偏差
skewness	歪度
kurtosis	尖度

2 提案方法

図 1 に提案モデルの概観を示す．提案モデルは発話音声からモダリティごとに特徴抽出し，それぞれ学習モデルを構築する．音響特徴量によって学習されたモデルを「音響モデル」，言語特徴量によって学習されたものを「言語モデル」として，2 つのモデルの出力を融合することで FTLD・ALS を鑑別する．

2.1 特徴量の定義

2.1.1 音響特徴量（384 種）

音響特徴量には INTERSPEECH 2009 Emotion Challenge[9] で用いられた 384 種の特徴量を用いる．これら特徴量は表 2 に示す音響特徴量に対して，表 3 の統計量を算出した値である．特徴抽出には音声解析ソフトウェア openSMILE[10] を使用した．

2.1.2 言語特徴量（22 種）

花井ら [7] が用いた言語特徴量 17 特徴および時間特徴量 3 特徴に「単位時間あたりの発話語数」と「単位時間あたりの異なり語数」を追加した計 22 特徴を言語特徴量と定義する．言語特徴量を抽出するには発話音声をテキストへ書き起こす必要があるが，人手による文字起こし作業は検査実施者の負担となると考えられる．そのため，本研究では音声解析システム Watson Speech to Text を用いることで発話音声の書き起こし作業を自動化している．発話音声をテキスト化した文字列に対して，形態素解析エンジン MeCab によって形態素解析を施した．

2.2 識別モデル

疾患の分類にはアンサンブル学習の一種である Random Forest（RF）を用いる．決定木の数は 100 で固定し，決定木を構成するための最大利用可能特徴数を ['sqrt', 'log2']，決定木の最大深度を [10, 20, 30, 40,50] と定めて，grid search によりハイパーパラメータチューニングをした．モダリティ X によって学習されたモデル $\mathrm{RF_X}$ と表記し，算出された事後確率 $Prob$ を出力する．

2.3 融合方法

音響モダリティと言語モダリティの融合方法について検討する．学習前にモダリティを結合する方法を「Early Fusion」，音響モデル $\mathrm{RF_{aco}}$ と言語モデル $\mathrm{RF_{lin}}$ の異なるモダリティから生成される学習器の出力結果を融合する方法を「Late Fusion」と呼ぶ．Late Fusion では訓練データから算出される F1 スコアを用いて融合する．音響モデルの F1 スコアを f_1^{aco}，言語モデルの F1 スコアを f_1^{lin} と記す．以下に従来手法である「Early Fusion」と，本稿にて考案する「Late Fusion 1」と「Late Fusion 2」を定義する．

2.3.1 Early Fusion

先行研究 [8] で用いられており，音声特徴 384 種と言語特徴 22 種を結合して合計 406 種の特徴量を用いて学習する．Early Fusion は次式 1 で表される．

$$Prob = \mathrm{RF}_{[\mathrm{aco,\ lin}]}. \tag{1}$$

2.3.2 Late Fusion 1

Late Fusion 1 は f_1^{aco} と f_1^{lin} の値を比較して値が大きいモデルの予測値を選択する．課題によって動的にモデルを切り替えることができる．Late Fusion 1 は次式 2 で定義される．

$$Prob = \begin{cases} \mathrm{RF_{aco}} & \text{if } \mathrm{f}_1^{lin} < \mathrm{f}_1^{aco}. \\ \mathrm{RF_{lin}} & \text{other.} \end{cases} \tag{2}$$

2.3.3 Late Fusion 2

$\mathrm{RF_{aco}}$ と $\mathrm{RF_{lin}}$ を加重平均によって融合する．音響モデルの重みを w_{aco}，言語モデルの重みを w_{lin} とする．それぞれ重みは f_1^{aco}，f_1^{lin} の値に基づいて算出する．分類クラスは 2 であるため，チャンスレートを 0.5 に設定し，重みは F1 スコアから 0.5 を引いた値とする．なお，一方のモダリティの F1 スコアが 0.5 未満の場合は他方のモダリティの事後確率を出力し，両方のモダリティの F1 スコアが 0.5 に至らない場合は単純平均によって融合する．Late Fusion 2 は次式 3 で定義される．

$$Prob = \frac{w_{aco} * \mathrm{RF_{aco}} + w_{lin} * \mathrm{RF_{lin}}}{w_{aco} + w_{lin}}. \tag{3}$$

$$w_{aco} = \begin{cases} \mathrm{f}_1^{aco} - 0.5 & \text{if } \mathrm{f}_1^{aco} > 0.5 \text{ and } \mathrm{f}_1^{lin} > 0.5. \\ 0 & \text{if } \mathrm{f}_1^{aco} \leq 0.5 \text{ and } \mathrm{f}_1^{lin} > 0.5. \\ 1 & \text{other.} \end{cases}$$

$$w_{lin} = \begin{cases} \mathrm{f}_1^{lin} - 0.5 & \text{if } \mathrm{f}_1^{aco} > 0.5 \text{ and } \mathrm{f}_1^{lin} > 0.5. \\ 0 & \text{if } \mathrm{f}_1^{aco} > 0.5 \text{ and } \mathrm{f}_1^{lin} \leq 0.5. \\ 1 & \text{other.} \end{cases}$$

表 4: 単語音読課題ごとの分類性能（FTLD・ALS / HC）

課題番号	単語課題	音響モデル	言語モデル	Early Fusion	Late Fusion 1	Late Fusion 2
課題 1	まど	0.70 ± 0.09	0.58 ± 0.10	**0.76 ± 0.09**	0.70 ± 0.10	0.67 ± 0.05
課題 2	パイプ	0.73 ± 0.07	0.63 ± 0.07	**0.78 ± 0.06**	0.68 ± 0.06	0.73 ± 0.04
課題 3	電話	0.68 ± 0.04	0.57 ± 0.08	**0.71 ± 0.08**	0.68 ± 0.08	0.67 ± 0.05
課題 4	なす	0.66 ± 0.11	0.57 ± 0.08	**0.68 ± 0.11**	0.63 ± 0.07	0.59 ± 0.15
課題 5	ばなな	0.75 ± 0.06	0.63 ± 0.07	**0.77 ± 0.08**	0.75 ± 0.07	0.67 ± 0.09
課題 6	33	0.68 ± 0.06	**0.71 ± 0.08**	**0.71 ± 0.05**	0.68 ± 0.08	**0.71 ± 0.06**
課題 7	ゆきだるま	0.70 ± 0.09	0.60 ± 0.10	**0.75 ± 0.12**	0.70 ± 0.10	0.70 ± 0.14
課題 8	25 パーセント	**0.79 ± 0.03**	0.71 ± 0.03	0.78 ± 0.06	0.75 ± 0.07	0.74 ± 0.06
課題 9	92 分の 1	0.67 ± 0.10	**0.73 ± 0.08**	0.70 ± 0.06	**0.73 ± 0.06**	0.71 ± 0.05
課題 10	とけいえんぴつ	**0.72 ± 0.07**	0.57 ± 0.04	0.71 ± 0.06	**0.72 ± 0.04**	0.69 ± 0.06

表 5: 文章音読課題ごとの分類性能（FTLD・ALS / HC）

課題番号	文章課題	音響モデル	言語モデル	Early Fusion	Late Fusion 1	Late Fusion 2
課題 11	電話が鳴っています	**0.72 ± 0.06**	0.62 ± 0.08	0.68 ± 0.10	**0.72 ± 0.08**	0.66 ± 0.08
課題 12	魚屋は元気でした	**0.72 ± 0.07**	0.64 ± 0.08	0.71 ± 0.10	0.68 ± 0.11	0.66 ± 0.12
課題 13	兄はまだ戻りません	0.75 ± 0.06	0.68 ± 0.05	**0.78 ± 0.08**	0.69 ± 0.09	0.73 ± 0.10
課題 14	日本高校野球連盟	0.73 ± 0.02	0.59 ± 0.09	**0.75 ± 0.11**	0.73 ± 0.09	0.69 ± 0.10
課題 15	だけどやっぱりでもはだめ	**0.73 ± 0.05**	0.70 ± 0.05	**0.73 ± 0.05**	0.71 ± 0.07	0.70 ± 0.05
課題 16	新しい甘酒を 5 本のひょうたんに入れなさい	0.73 ± 0.04	0.64 ± 0.03	**0.77 ± 0.06**	0.66 ± 0.08	0.67 ± 0.09

3 対象疾患と実験データセット

3.1 対象疾患

本研究は非専門医の FTLD 診断を支援するスクリーニングツールの開発を目的とする．一方で FTLD は希少疾患でありデータ数が少ないため，実験では FTLD と関連性が指摘されている ALS を疾患群に含め，FTLD+ALS 群を対象疾患群とする．

3.1.1 前頭側頭葉変性症（FTLD）

FTLD は大脳の前頭葉や側頭葉を中心に神経細胞の変性・脱落によって，言葉の意味の理解や物の名前などの知識が失われる「語義失語」や発話量の減少などの症状が見られ，行動障害，認知機能障害などが緩徐に進行する神経変性疾患である．FTLD は脳の病変部位により，意味性認知症（Semantic Dementia: SD），進行性非流暢性失語症（Progressive Non-Fluent Aphasia: PNFA），行動障害型前頭側頭型認知症（behavioral variant Frontotemporal Dementia: bvFTD）の 3 つの臨床疾患に分けられている．主として初老期に発症し，症状には人格変化や社会行動の乱れが現れるため，診断が遅れる症例や社会的に問題になる症例が報告されている．

3.1.2 筋萎縮性側索硬化症（ALS）

筋萎縮性側索硬化症（Amyotrophic Lateral Sclerosis: ALS）は運動神経細胞に障害を与える神経変性疾患である．全身の筋肉萎縮により，四肢の筋力低下や構音障害などの症状が見られる．また ALS は認知症を発症した場合，前頭葉機能が低下し FTLD と同一のタンパクの異常蓄積が見られることから，FTLD との連続性が指摘されている [11].

3.2 実験協力者

表 1 に実験協力者の内訳を示す．実験には対象疾患患者に加え，HC を含めた 168 人（年齢 32〜82 歳，男性 76 人，女性 92 人）が参加し，名古屋大学医学部附属病院，名古屋大学医学部保健学科大幸キャンパス並びに大阪大学医学部附属病院にて音声データを収集した．HC は MMSE スコア ≥26 かつ，ACE-R スコア ≥89 の条件を満たした者と定義した．また音響特徴は性別や年齢に影響されるため，HC は疾患群の年齢・性別に合うように統制した．なお，本研究は研究内容および方法について参加した各機関における倫理審査委員会の承認を得ている．

3.3 音読課題

実験協力者に対して，音読課題を実施した．音読課題とはモニターに表示された単語又は文章を被験者が見て音読する課題である．課題は WAB 失語症検査 [12] の「復唱課題」で用いられている単語および文章に「とけいえんぴつ」を追加した計 16 課題を使用する．音読課題の実施順序は復唱課題と同じ順番とする．表 4 に単語音読課題，表 5 に文章音読課題の一覧を示す．

4 性能評価実験

FTLD・ALS の鑑別モデルについて，性能評価を比較・検証する．実験では各課題について音響モダリティのみ，言語モダリティのみで学習したモデルと，Early Fusion，Late Fusion 1，Late Fusion 2 についてそれぞれ性能評価を算出した．汎化性能の評価には層化抽出法による 5 分割交差検証を使用し，評価指標には F1 スコアを用いる．なお，使用したデータセットには一部欠損値が含まれているため，訓練データの欠損値は疾患ごとの平均値を代入し，テストデータの欠損値には全訓練データの平均値を代入することで補完した．また訓練データは特徴量ごとに標準化し，テストデータは訓練データの平均および標準偏差に基づいて標準化している．

5 結果と考察

表 4，5 に実験の結果を示す．課題ごとに 5 つのモデルの中で最も優れた F1 スコアを太字で表している．音響モデルと言語モデルの F1 スコアを比較すると課題 6，9 を除いた全ての課題で音響モデルが優れている．音読課題はいずれも発話内容が決まっているため，言語モデルが扱う特徴量には認知機能の差が現れにくい．特に単語課題ではこの傾向が顕著である．また FTLD・ALS に

表 6: Late Fusion 2 の融合時の重み

課題番号	w_{aco}	w_{lin}
課題 1	**0.26 ± 0.02**	0.15 ± 0.03
課題 2	**0.26 ± 0.02**	0.22 ± 0.01
課題 3	**0.20 ± 0.03**	0.10 ± 0.01
課題 4	**0.14 ± 0.05**	0.10 ± 0.01
課題 5	**0.24 ± 0.02**	0.15 ± 0.02
課題 6	0.17 ± 0.02	**0.18 ± 0.03**
課題 7	**0.21 ± 0.02**	0.12 ± 0.03
課題 8	**0.23 ± 0.01**	0.19 ± 0.04
課題 9	0.19 ± 0.02	**0.24 ± 0.02**
課題 10	**0.18 ± 0.02**	0.13 ± 0.02
課題 11	**0.16 ± 0.02**	0.11 ± 0.03
課題 12	**0.16 ± 0.03**	0.15 ± 0.05
課題 13	0.20 ± 0.03	**0.23 ± 0.02**
課題 14	**0.26 ± 0.01**	0.13 ± 0.05
課題 15	**0.19 ± 0.02**	**0.19 ± 0.02**
課題 16	**0.18 ± 0.02**	**0.18 ± 0.04**

みられる構音障害は主に音響特徴として抽出される．以上の理由から音響モデルの性能が言語モデルに比べて高いと考える．

Late Fusion 1 に着目する．Late Fusion 1 では訓練データから算出した F1 スコアの値によって音響モデルと言語モデルを動的に選択している．単語課題の場合，言語モデルに比べて音響モデルの性能が高くなるため，Late Fusion 1 において音響モデルが選択されやすい．文章課題の場合，単語課題に比べて言語モデルの F1 スコアが高いが，音響モデルには劣る．そのため単語課題と同様に音響モデルが選択されやすい．一方で，Early Fusion は音響と言語の特徴量を結合して学習するため，それぞれのモダリティから FTLD・ALS 判別に寄与する特徴量を重要視する．そのため Early Fusion が音響を重視する Late Fusion 1 よりも性能が上回る課題数が多いと考える．

次に Late Fusion 2 に着目する．Late Fusion 2 の F1 スコアは課題 6 でのみ最良となった．表 6 に各課題における音響モデルの重み w_{aco} と言語モデルの重み w_{lin} を示す．重みは課題 6，9，13 のときに w_{lin} が w_{aco} より大きい．しかし表 4 および表 5 の結果と比較すると，必ずしも重みが大きいモダリティのモデルが他方のモデルの F1 スコアよりも高いとは言えない．重みは訓練データから算出された F1 スコアを用いているため，訓練データの分布とテストデータの分布が異なるのではないかと推測する．

音響モダリティと言語モダリティの混合方法を検討したが，Late Fusion 1 と Late Fusion 2 ともに従来手法の Early Fusion に劣る結果となった．しかしながら，課題 6，9 において言語モデルおよび Late Fusion の有効性が示唆された．今後は言語モダリティの性能の向上が期待される自発話会話や導入発話の回答音声を解析し，音響モダリティと言語モダリティの融合方法について検討する．

6 おわりに

本研究では，モダリティの違いを考慮した認知症スクリーニングシステム開発を目指し，音声と言語のモダリティ混合による FLTD・ALS 鑑別モデルを提案した．16 種の音読課題の回答音声から音響モダリティと言語モダリティを抽出し，モダリティごとに生成したモデルの融合方法を考案した．その結果，Late Fusion 1 や Late Fusion 2 は従来手法である Early Fusion に劣る結果となったものの，言語モデルの性能が高い一部の課題において有効性が示唆された．今後は FTLD における語義失語などの症状をふまえ，導入発話や自発話課題の回答音声を解析し，音響モダリティと言語モダリティの融合手法について検討する．

謝辞

本研究は，一部，文部科学省科学研究費補助金（課題番号 JP19H01137）の助成により行われた．

参考文献

[1] 厚生労働省：認知症施策推進総合戦略 (新オレンジプラン), 認知症地域支援推進員 (概要). http://www. mhlw. go. jp/stf/seisakunitsuite/bunya/0000064084. html,(参照日 2022 年 6 月 5 日) (2015).

[2] Mendez, M. F. and Cummings, J. L.: *Dementia: a clinical approach*, Butterworth-Heinemann (2003).

[3] Snowden, J. S.: Frontotemporal lobar degeneration: Frontotemporal dementia, progressive aphaisia, semantic dementia, *Clinical neurology and neurosurgery monographs* (1996).

[4] Folstein, M. F., Folstein, S. E. and McHugh, P. R.:“Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician, *Journal of psychiatric research*, Vol. 12, No. 3, pp. 189–198 (1975).

[5] 加藤伸司：改訂長谷川式簡易知能評価スケール (HDS-R) の作成, 老年精神医学雑誌, Vol. 2, pp. 1339–1347 (1991).

[6] Kato, S., Homma, A. and Sakuma, T.: Easy screening for mild alzheimer's disease and mild cognitive impairment from elderly speech, *Current Alzheimer Research*, Vol. 15, No. 2, pp. 104–110 (2018).

[7] 花井俊哉, 加藤昇平, 坂口巧一, 佐久間拓人, 大嶽れい子, 桝田道人, 渡辺宏久：認知課題遂行時の発話特徴を用いた認知症希少疾患の簡易検出, 電子情報通信学会論文誌 D, Vol. 104, No. 4, pp. 198–206 (2021).

[8] 坂口巧一, 加藤昇平, 花井俊哉, 佐久間拓人, 大嶽れい子, 桝田道人, 渡辺宏久：音読課題音声からの FTLD・ALS 簡易検出モデル, 情報処理学会第 83 回全国大会, Vol. 1, p. 02 (2021).

[9] Kockmann, M., Burget, L. and Cernockỳ, J.: Brno University of Technology system for Interspeech 2009 emotion challenge., in *Interspeech*, pp. 348–351 (2009).

[10] Eyben, F. and Schuller, B.: openSMILE:) The Munich open-source large-scale multimedia feature extractor, *ACM SIGMultimedia Records*, Vol. 6, No. 4, pp. 4–13 (2015).

[11] Riku, Y., Watanabe, H., Yoshida, M., Tatsumi, S., Mimuro, M., Iwasaki, Y., Katsuno, M., Iguchi, Y., Masuda, M., Senda, J., *et al.*: Lower Motor Neuron Involvement in TAR DNA-Binding Protein of 43 kDa–Related Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis, *JAMA neurology*, Vol. 71, No. 2, pp. 172–179 (2014).

[12] 杉下守弘, 亀和田文子：WAB 失語症検査, 失語症研究, Vol. 7, No. 3, pp. 222–226 (1987).