1H2-NFC3b-12

音声を用いた認知機能障害の早期スクリーニングをめざして

–高齢者音声韻律特徴を用いた予備的研究–

Toward Early Detection of Cognitive Impairment in Elderly Using Speech Prosody

加藤昇平[*1] 半谷幸寛[*1] 小林朗子[*2] 小島敏昭[*2] 伊藤英則[*1] 本間昭[*3]
Shohei Kato Sachio Hanya Akiko Kobayashi Toshiaki Kojima Hidenori Itoh Akira Homma

[*1]名古屋工業大学 Nagoya Institute of Technology
[*2]イフコム Ifcom Co., Ltd.
[*3]認知症介護研究・研修東京センター Tokyo Dementia Care Research and Training Center

This paper presents a new trial approach to early detection of cognitive impairment in the elderly with the use of speech sound analysis and multivariate statistical technique. In this paper, we focus on the prosodic features from speech sound. Japanese 115 subjects (32 males and 83 females between ages of 38 and 99) participated in this study. We collected speech sound in the a few segments of dialogue of HDS-R examination. The segments corresponds to speech sound that is answering for questions on time orientation and number backward count. Firstly, 130 prosodic features have been extracted from each of the speech sounds. These prosodic features consist of spectral and pitch features (53), formant features (56), intensity features (19), and speech rate and response time (2). Secondly, these features are refined by principal component analysis and/or feature selection. Lastly, we have calculated speech prosody-based cognitive impairment rating (SPCIR) by multiple linear regression analysis. The results indicated that there is moderately significant correlation between HDS-R score and synthesis of several selected prosodic features. Consequently, adjusted coefficient of determination $\bar{R}^2 = 0.50$ suggests that prosody-based speech sound analysis has possibility to screen the elderly with cognitive impairment.

1. はじめに

日本国における超高齢化社会の進展は目覚ましく，最近の推計結果では国内の認知症者数は約 200 万人以上と言われており，2015 年には約 302 万人に倍増することが報告されている [Awata 09]．こうした中，平成 20 年 7 月，厚生労働省において「認知症の医療と生活の質を高める緊急プロジェクト」が取りまとめられ，今後の認知症対策の 1 つの柱として認知症の早期診断の重要性が掲げられている．

現在，認知症のスクリーニングは，HDS-R（改訂長谷川式簡易知能評価スケール）[Katoh 91]，MMSE（Mini-Mental State Examination）[Folstein 75]，CDR（Clinical Dementia Rating）[Morris 93] などが，fMRI，FDG-PET，CSF バイオマーカーなどの神経生理学に基づくテストと同様に広く用いられている．これらは一定のトレーニングを受けた医師，あるいは臨床心理士などにより，主として医療機関において実施されている．しかしながら，日常の外来診療場面では，HDS-R などの簡易検査であっても，5〜20 分程度の時間を要し，他の外来患者の診療に支障をきたすとの指摘もあり，医師の負担の軽減が重要になると考えられる．そこで，さらに簡便な非侵襲的，かつ，従来のツールと同等以上の性能を有するツールが開発されれば，より広範にスクリーニングを実施することが可能となり，認知症の早期診断に資することが可能になる．

本研究では，高齢者の発話音声に着目する．Taler ら [Taler 08] は，非常に軽度のアルツハイマー病患者を対象に文法および感情韻律上に障害が認められることを報告している．また，Hoyte ら [Hoyte 09] は，高齢者の発語の韻律特徴は，統語構造を検出する上で有用であることを示している．これらの報告は，韻律特徴の解析を，認知症のスクリーニング検査として用いることができる可能性を示唆している．本稿では，患者の日常会話音声から音声韻律特徴を解析することで，認知症の早期診断を実現することを目的とする．これにより，かかりつけ医の日常の診療時間に対する影響はほぼ皆無となり，認知症の早期発見が極めて容易になる．本稿では，同目的の予備的研究として，高齢者の HDS-R スコアと発話音声の韻律特徴との相関性を明らかにする．そして，健常者（NL）および認知機能障害患者（CI）の弁別における音声韻律特徴の有効性について議論する．

連絡先: 加藤昇平，名古屋工業大学大学院工学研究科情報工学専攻，〒 466- 8555 名古屋市昭和区御器所町，052-735-5625，shohey@nitech.ac.jp

2. 認知機能の指標と実験参加者

本研究では，高齢者の認知機能障害の指標として，国内の医療現場で広く用いられている，改訂長谷川式簡易知能評価スケール（HDS-R）[Katoh 91] を採用する．同スケールは口頭による質疑応答形式のテストとして実施されるため，実験に参加した高齢者からテストにおける音声会話を記録した．本稿では，質疑応答の中から「見当識」と「数字逆唱」についての２問の回答音声を収集した．これに追加して，高齢者が，出身地，子供のころの遊び，学生時代，の３つのテーマについて雑談したものから任意の発話音声の冒頭１フレーズについても収集した．

実験には 115 名の高齢者（年齢 38〜99 歳，男性 32 名，女性 83 名）が参加した．収集された音声資料の総数は 319 である．表 1 に内訳を示す．そして，被験者の HDS-R スコアに応じてそれぞれの事例を健常（NL）（HDS-R スコア 30〜24）と認知機能障害（CI）（HDS-R スコア 23〜11）の 2 クラスに分類（NL=205 データ，CI=114 データ）し，分析を行った．

3. 音声韻律特徴

音声は，3 つの要素（韻律，音質，音韻）から成り立つ．この中でも，韻律的特徴が人間の感情表現（例えば，文献 [Cowie 01, Scherer 03, Cho 09]）や認知機能障害（例えば，文献 [Taler 07, Taler 08]）などを特定するために重要な非言語情報となり得ることが明らかにされている．本研究では，以下に記述する 130 種の音声韻律特徴を抽出する．特徴抽出の際には，量子化

表 1: A Breakdown List of 319 Speech Data (N=115)

Age	30's	40's	50's	60's	70's	80's	90's	Total
Male	3 (1)	0 (0)	15 (5)	32 (11)	21 (7)	12 (5)	7 (3)	90 (32)
Female	0 (0)	20 (7)	45 (15)	24 (8)	28 (10)	87 (33)	25 (10)	229 (83)
Subtotal	3 (1)	20 (7)	60 (20)	56 (19)	49 (17)	99 (38)	32 (13)	319 (115)

Value in bracket means the number of subjects.

ビット数 16 およびサンプリング周波数 44.1[KHz] のデジタル音声を用い，短時間分析におけるフレーム長を 23[msec]，フレーム周期 11[msec] とし，窓関数として Hamming 窓（1024 ポイント）を使用した．

3.1 スペクトルとピッチ

ここでは，音声の高さに関係するピッチ構造を反映させるために，基本周波数と基本周波数の n 倍の周波数を持つ n 次高調波成分から得られる特徴量を以下に示す 53 個抽出する．ただし，基本周波数の時間変化の振幅とは，1 事例の基本周波数のデータ列の上位 25 %と下位 25 %の値を無視したときの最大値と最小値の幅とする．

1.-7. 発音開始直後 t 秒間の基本周波数の時間変化の振幅（$t = 0.05, 0.10, \cdots, 0.35$）

8. 周波数重心（各高調波成分のパワー値を重みとする周波数の重みつき平均）

9. 全高調波成分のパワー値の合計に対する基音成分のパワー値の割合

10.-48. 全高調波成分のパワー値の合計に対する基音から n 次までの高調波成分のパワー値の合計の割合（$n = 2, 3, \cdots, 40$）

49. 奇数次の高調波成分（基音含む）と偶数次の高調波成分とのパワー値の合計の比

50.-53. 基本周波数の標準偏差，平均，最大，最小値

3.2 フォルマント

ここでは，音声の特徴を表すフォルマント構造を反映させるために，フォルマント周波数とフォルマント帯域幅を特徴量として以下に示す 56 個抽出する．

54.-57. 第 n フォルマント周波数の標準偏差（$n = 1, \cdots, 4$）

58.-61. 第 n フォルマント周波数の平均値（$n = 1, \cdots, 4$）

62.-65. 第 n フォルマント周波数の最大値（$n = 1, \cdots, 4$）

66.-69. 第 n フォルマント周波数の最小値（$n = 1, \cdots, 4$）

70.-73. 第 n フォルマント周波数の中央値（$n = 1, \cdots, 4$）

74.-77. 第 n フォルマント周波数の最大値と最小値の差（$n = 1, \cdots, 4$）

78.-81. 第 n フォルマント周波数の線形近似直線の傾き（$n = 1, \cdots, 4$）

82.-85. 第 n フォルマント帯域幅の標準偏差（$n = 1, \cdots, 4$）

86.-89. 第 n フォルマント帯域幅の平均値（$n = 1, \cdots, 4$）

90.-93. 第 n フォルマント帯域幅の最大値（$n = 1, \cdots, 4$）

94.-97. 第 n フォルマント帯域幅の最小値（$n = 1, \cdots, 4$）

98.-101. 第 n フォルマント帯域幅の中央値（$n = 1, \cdots, 4$）

102.-105. 第 n フォルマント帯域幅の最大値と最小値の差（$n = 1, \cdots, 4$）

106.-109. 第 n フォルマント帯域幅の線形近似直線の傾き（$n = 1, \cdots, 4$）

3.3 エネルギー

ここでは，音声の大きさに関係する振幅構造を反映させるために，短時間パワーとその包絡線から得られる特徴量を以下に示す 19 個抽出する．

110. パワー包絡線の線形最小二乗法による近似直線の傾き

111.-117. 発音開始直後 t 秒間のパワー包絡線の微分係数の中央値（$t = 0.05, 0.10, \cdots, 0.35$）

118.-124. 最大パワー値と発音開始から t 秒後のときのパワー値の比（$t = 0.05, 0.10, \cdots, 0.35$）

125.-128. 短時間パワーの標準偏差，平均，最大，最小値

3.4 Speech Rate and Response Time

ここでは，時間構造を反映させるために，被験者が話す速さ，ならびに，質問に回答するまでの反応時間の 2 つの特徴量を抽出する．

129. 1 モーラあたりの発話継続時間

130. 返答までの反応時間

4. 特徴選択

本研究では，高齢者の音声データから音声韻律特徴を抽出し，解析することで認知機能障害（CI）と健常（NL）を判別することを目的としている．しかしながら，解析を行なう際に，データから抽出した特徴量が多すぎると，その中には認知機能障害の判別に無関係な特徴量が含まれる可能性があり，モデルの構築や判別の精度に悪影響を与えることが考えられる．また，特徴量が多すぎるとモデルが複雑になりすぎたり，計算コストが高くなる短所もある．そこで本稿では，前章で述べた音声韻律特徴に特徴選択を行なう．

特徴選択の手法としては，科学的理論や事前の知識によって適当な変数を指定する変数指定法や，全ての変数の組合せを計算し，最良と思われるものを選択する総当たり法，一定の規則にしたがって変数を逐次選択していく逐次選択法などがあげられる．現在のところ，高齢者の認知機能障害と因果関係の高い音声特徴は特定されておらず，特徴選択として有用な理論や事前の知識は存在しない．また，抽出した特徴量のすべての組合せを計算することは計算コストが高くなる．そのため，一般的に多用されている逐次選択法としてフォワードステップワイズ法（FSW）[Draper 98] を用いて特徴選択を行なう．フォワードステップワイズ法の特徴選択規準としては，赤池情報量規準（AIC）[Akaike 74] を採用する．本稿では，特徴選択の事前に主成分分析（PCA）による変数合成を施す手法も併せて検討する．

次節では，AIC 規準を用いた FSW 法（FSW-AIC），事前に PCA による変数合成を施した FSW 法（PCA-FSW-AIC），ならびに，全特徴量を強制投入する手法（FE，特徴選択なし）の 3 つの方法で HDS-R スコアと音声韻律特徴の相関を分析する．

表 2: Correlation between SPCIR and HDS-R by Multiple Linear Regression

	SPCIR_{FE}	$\text{SPCIR}_{\text{FSW-AIC}}$	$\text{SPCIR}_{\text{PCA-FSW-AIC}}$
# of regressors	130	19	55
R	0.78	0.67	0.77
$\bar{R}^2$	0.37	0.41	0.50
S.E.	4.57	4.43	4.08

表 3: Dominant Regressors for Estimate of HDS-R

Method		dominant regressors
SPCIR_{FE}		130 regressors in total
	***	F129
	**	F110
	*	F57, F33, F78
$\text{SPCIR}_{\text{FSW-AIC}}$		19 regressors in total
	***	F129, F128
	**	F118, F130, F57, F8, F101, F59
	*	F110, F72, F69, F73
$\text{SPCIR}_{\text{PCA-FSW-FE}}$		55 regressors in total
	***	PC2, PC7, PC12, PC4, PC26, PC52, PC54
	**	PC34, PC3, PC30, PC9, PC77, PC15, PC61, PC115
	*	PC22, PC40, PC13, PC31, PC103, PC14, PC1, PC46, PC129, PC100

***: with significance level of 0.001, **: with significance level of 0.01
*: with significance level of 0.05

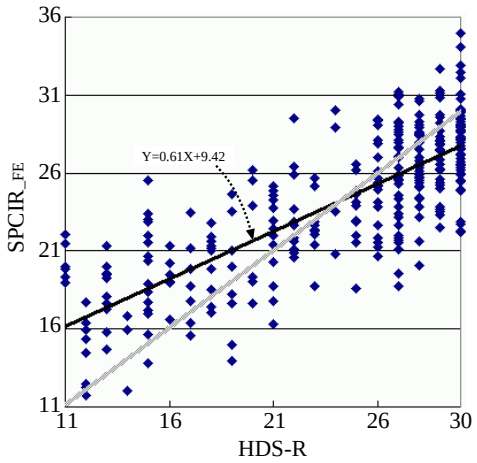

図 1: Scatter plot of HDS-R and SPCIR_{FE} ($\bar{R}^2 = 0.37$).

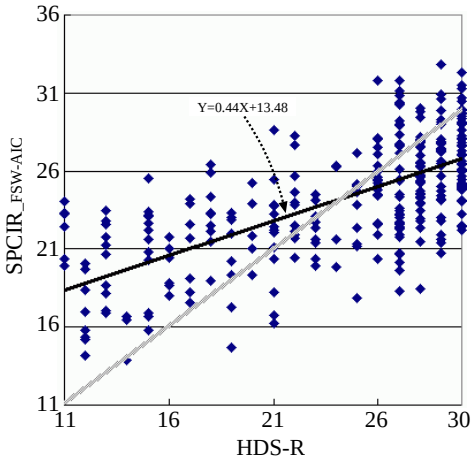

図 2: Scatter plot of HDS-R and $\text{SPCIR}_{\text{FSW-AIC}}$ ($\bar{R}^2 = 0.41$).

5. 音声韻律に基づく認知機能障害評定

本節では，高齢者の発話音声 319 データ（N=115）における HDS-R スコアと音声韻律特徴の相関性について述べる．前節で述べた特徴選択手法により選択・合成された音声特徴を説明変数とし，高齢者の HDS-R スコアを目標属性として重回帰分析を行うことにより，**音声韻律に基づく認知機能障害評定（SPCIR: speech prosody-based cognitive impairment rating）**を導出した．SPCIR_{FE}，$\text{SPCIR}_{\text{FSW-AIC}}$，，$\text{SPCIR}_{\text{PCA-FSW-AIC}}$ は，それぞれ，FE，FSW-AIC，PCA-FSW-AIC の変数選択法に基づく評定である．表 2 にこれらの評定による相関性の結果を示す．図 1-3 に HDS-R スコアと SPCIR の散布図を示す．表 3 に，それぞれの特徴選択を用いた重回帰分析の結果，有意と判断された特徴量の内訳を示す．

SPCIR_{FE}： 一見して HDS-R スコアとの高い相関性を有する（$R = 0.78$）と見えるものの，補正済み決定係数が $\bar{R}^2 = 0.37$ と落ち込んでいる．この手法で有意な回帰係数を持つ変数もあまり見つかっていない．これは全特徴を用いた強制投入により過学習が生じた結果と考えられる．

$\text{SPCIR}_{\text{FSW-AIC}}$： AIC 規準を用た変数選択により過学習を回避でき，有意な回帰係数を持つ変数が増加している．しかしながら，HDS-R スコアとの相関（$R = 0.67$）および補正済み決定係数が $\bar{R}^2 = 0.41$ と十分とは言えない結果となった．この手法では，AIC 規準が持つモデルの複雑さに関するペナルティ項が働いたことにより 19 個の変数しか選ばれなかったが，残りの 119 変数の中にはまだ HDS-R スコアの推定に有用な変数が含まれていると考えられる．

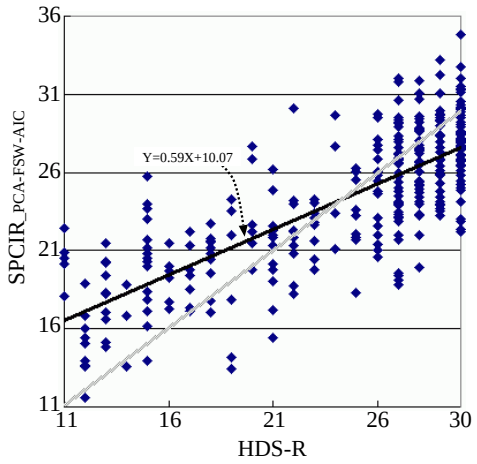

図 3: Scatter plot of HDS-R and $\mathrm{SPCIR}_{\text{PCA-FSW-AIC}}$ ($\bar{R}^2 = 0.50$).

$\mathrm{SPCIR}_{\text{PCA-FSW-AIC}}$： AIC 規準を用いた特徴選択の事前に PCA による変数合成を行うことで上記 2 手法の問題点が改善されている．この手法では，130 個の全特徴を適切に合成した 55 個の主成分が回帰式に採用されている．表 3 の結果から，多くの変数が有意な回帰係数を持ち，高い固有値を持つ主成分（例えば PC2, PC7, PC4）のみならず，PC77，PC115，PC103 などの低い固有値を持つ主成分も HDS-R スコアの推定に有用であることが示された．

以上の結果と図 3 の散布図から，音声韻律特徴を用いた認知機能障害の評定 $\mathrm{SPCIR}_{\text{PCA-FSW-AIC}}$ は HDS-R スコアとの間にある程度強い相関（$R = 0.77$）を持つことが示唆された．本分析結果が示した補正済み決定係数 $\bar{R}^2 = 0.50$ は，韻律特徴に基づく高齢者の発話音声解析の認知機能障害のスクリーニングへの応用可能性として有意義な結果である．

6. おわりに

本研究では，高齢者の認知機能障害のスクリーニング手法開発のための新しい試みとして，発話音声の韻律特徴解析手法および高次統計量解析を融合させたアプローチを提案した．本稿では，臨床試験の予備的研究として，115 名の高齢者から 319 の音声資料を収集し，130 種の音声韻律特徴を抽出 · 選択 · 合成することで，音声による認知機能障害評定が導出できることを実験的に示した．同評定が医療現場で広く用いられている HDS-R（改訂長谷川式簡易知能評価スケール）スコアとの間に比較的有意な相関を持つことも確認された．その結果提案手法の認知機能障害のスクリーニング手法としての応用可能性を確認した．

今後の課題としては，高齢者データを増加することによる分析 · 推定性能の向上，fNIRS などの非侵襲な神経生理学的脳機能計測を組合せることによるマルチモーダルなスクリーニング技術の開発，ならびに，臨床試験を経ることで次世代の認知症のスクリーニングツールを開発したいと考えている．

謝辞

本研究は，一部，科学技術振興機構（JST）先端計測分析技術 · 機器開発プロジェクトの資金により行われた．

参考文献

[Akaike 74] Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723 (1974)

[Awata 09] Awata, S.: Roll of the dementia medical center in the community, in *Japanese Journal of Geriatrics*, Vol. 46, pp. 203–206 (2009), (in Japanese)

[Cho 09] Cho, J., Kato, S., and Itoh, H.: Comparison of Sensibilities of Japanese and Koreans in Recognizing Emotions from Speech by using Bayesian Networks, in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2945–2950 (2009)

[Cowie 01] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G.: Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, Vol. 18, No. 1, pp. 32–80 (2001)

[Draper 98] Draper, N. and Smith, H.: *Applied Regression Analysis (3rd edition)*, John Wiley & Sons (1998)

[Folstein 75] 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, *J. Psychiat. Res*, Vol. 12, No. 3, pp. 189–198 (1975)

[Hoyte 09] Hoyte, K., Brownell, H., and Wingfield, A.: Components of Speech Prosody and their Use in Detection of Syntactic Structure by Older Adults, *Experimental Aging Research*, Vol. 35, No. 1, pp. 129–151 (2009)

[Katoh 91] Katoh, S., Simogaki, H., Onodera, A., Ueda, H., Oikawa, K., Ikeda, K., Kosaka, K., Imai, Y., and Hasegawa, K.: Development of the revised version of Hasegawa's Dementia Scale (HDS-R), *Japanese Journal of Geriatric Psychiatry*, Vol. 2, No. 11, pp. 1339–1347 (1991), (in Japanese)

[Morris 93] Morris, J. C.: The Clinical Dementia Rating (CDR): Current version and scoring rules, *Neurology*, Vol. 43, No. 11, pp. 2412–2414 (1993)

[Scherer 03] Scherer, K. R., Johnstone, T., and Klasmeyer, G.: *Vocal expression of emotion*, R. J. Davidson, H. Goldsmith, K. R. Scherer eds., Handbook of the Affective Sciences (pp. 433–456), Oxford University Press (2003)

[Taler 07] Taler, V. and Phillips, N.: Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review, *Journal of Clinical and Experimental Neuropsychology*, Vol. 30, No. 5, pp. 501–556 (2007)

[Taler 08] Taler, V., Baum, S. R., Chertkow, H., and Saumier, D.: Comprehension of grammatical and emotional prosody is impaired in Alzheimer's disease, *Neuropsychology*, Vol. 22, No. 2, pp. 188–195 (2008)