

コンピュータエージェントによる 質問への応答からの 認知症の自動検出

宇城毅儀¹, 田中宏季¹, 足立浩祥², 數井裕光³, 池田学⁴,
工藤喬², 中村哲¹



- 1 奈良先端科学技術大学院大学 情報科学研究科
- 2 大阪大学 キャンパスライフ健康支援センター
- 3 高知大学 医学部 神経精神科学教室
- 4 大阪大学大学院 医学系研究科

認知症対策の重要性

- 患者数の増加

(2012年) (2025年)
日本における患者数：約462万人 → 約675万人

(2015年) (2050年)
世界における患者数：約4700万人 → 約1億3000万人

- 患者が与える周囲への影響

- 未確立である治療法

認知症の早期発見が重要

音声言語に着目した認知症検出

認知機能の低下が音声言語に反映

[Taler+, 2008][Fraser+, 2016][池田, 2015]



音声言語に着目した先行研究

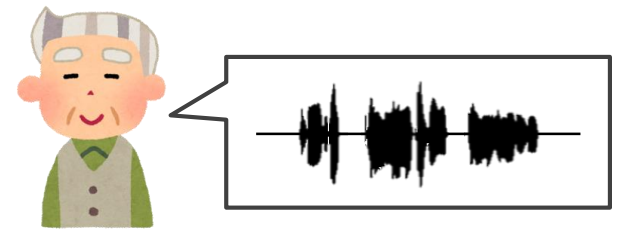
- 神経心理学的検査中の発話からの認知症検出 [Roark+, 2014]
- 質問応答中の音声からの認知症検出 [Kato+, 2018]
- エージェントへの質問応答からの認知症検出 [Tanaka+, 2017]

質問応答中の音声からの認知症検出 [Kato+, 2018]

質問応答中の音声

質問

出生地, 卒業した小学校名,
日時見当識, 3桁の数字の逆唱

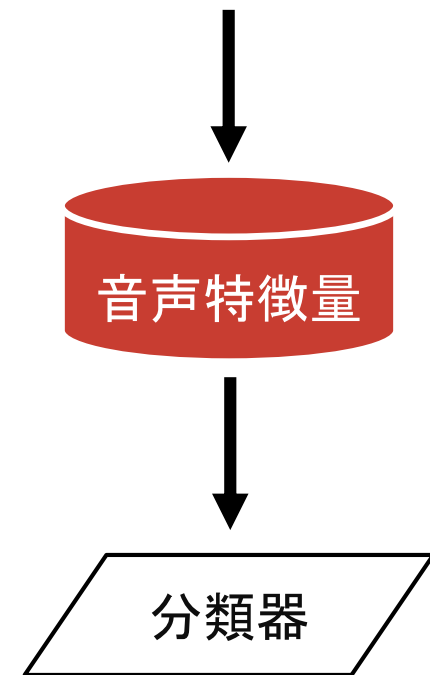


音声特徴量の抽出

回答するまでの反応時間, ピッチ,
フォルマント, 声の大きさ など

分類正解率 (認知症: 91名/非認知症: 91名)

出生地: 67.2%, 小学校名: 74.7%,
日時見当識: 89.5%, 数字の逆唱: 78.1%



認知症検出に有効な特徴量

認知症患者の反応時間の遅れ [Pirozzolo+, 1981]



『質問を投げかけてから応答するまでの時間』
を特徴量の1つとして抽出 [Kato+, 2018][Tanaka+, 2017]



『質問を投げかけてから応答するまでの時間』
が最も有効な特徴量 [Tanaka+, 2017]

先行研究の課題

先行研究の 提案

(音声言語から認知症を検出するために)
有効な特徴量や検出手法

先行研究の 問題点

課題が同じパターン

→ 同じ課題に対する『慣れ』の影響

[Beglinger+, 2005]

→ 繰り返しの利用に不向き

繰り返しの利用を想定した
認知症の自動検出

研究目的と提案手法

目的

繰り返しの利用を想定した
認知症の自動検出

提案

質問をランダムで5問提示する
エージェントシステムによる自動検出

ランダムで出題

毎回同じ質問の提示を回避

→ 繰り返しの利用による『慣れ』に対応

用意した質問

人の見当識に係る質問（1問）

- 家族について

自己認識に係る質問（4問）

- ストレス, 利き手, 睡眠, 食欲について
- OSCE（問診技法などの評価テスト）に基づいて作成

本人の関心に係る質問（2問）

- 趣味, 好きな歌について

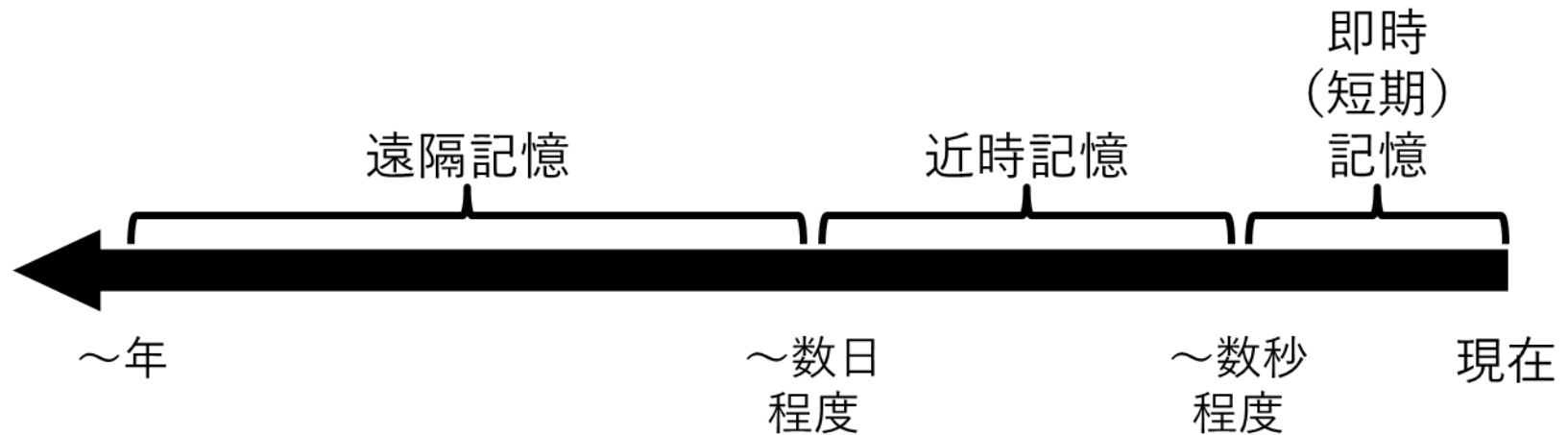
遠隔記憶に係る質問（3問）

- 過去の有名人について

神経心理学的検査で使用される質問（3問）

- 常識, 時の見当識

時間による記憶の区分



認知症患者の特徴

- 近時記憶の障害が目立つ
- 遠隔記憶は保たれていることが多い

神経心理学的検査（スクリーニング検査）

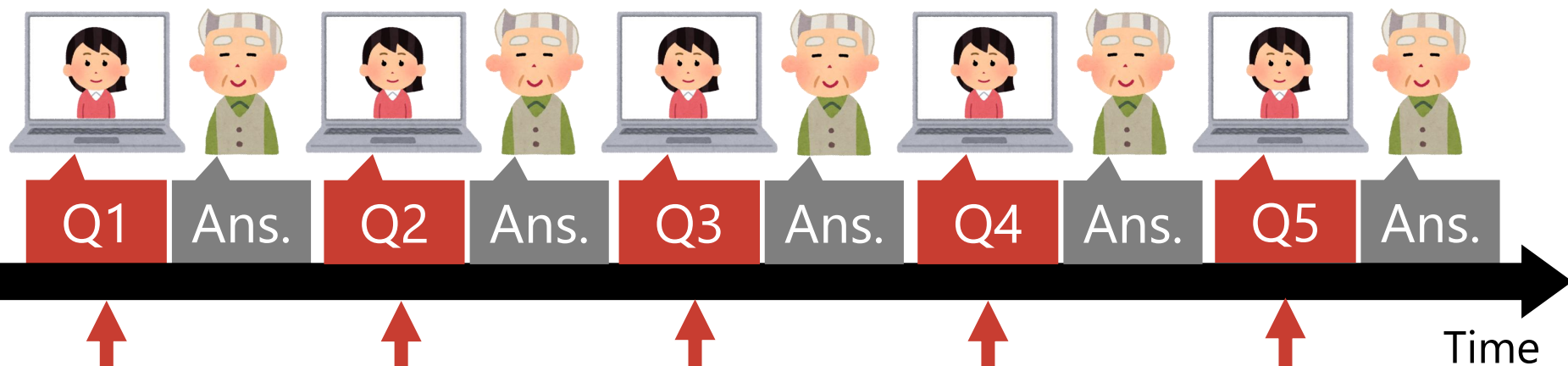
- 短時間で簡便に行える検査
- 先行研究の課題（質問）として用いられている
- 同じ課題のパターン

Mini-Mental State Examination（MMSE）

- 30点満点，カットオフ値：23/24，所要時間：約10分
- 面接方式
- 認知機能を評価

（課題例）100から7を順に引き算してください
（鉛筆を見せながら）これは何ですか？

質疑の流れ



質問例 (合計13問)

- 趣味は何ですか？
- 長嶋茂雄さんについて知っていることを教えてください

エージェントが13問の質問から**5問をランダム**で出題

エージェントシステム



- 先行研究 [Tanaka+, 2017] のシステムを使用
- 発話速度を低下
- 発話内容に字幕を付与

自動化の処理の流れ

応答音声



音声区間検出 (VAD)

- adintool



特徴量抽出

- 音声分析 : Snack Sound Toolkit
- 音声認識 : Julius
- 形態素解析 : MeCab toolkit



分類



認知症の
疑いなし

特徴量の抽出

音声特徴量

Utterance length: 回答（発話）の長さ

Gap: 質問を投げかけてから応答するまでの時間

Pause: (1秒以上の) 発話と発話の間の長さ

Power: 声の大きさ

基本周波数 (f_0): 声の高さ

言語特徴量

Speech rate: 発話速度, トークンの数, フィラーの数
品詞 (名詞, 動詞, 形容詞, 副詞) の数

研究協力者

● DSM-IV-TRに基づいて診断

グループ	N	年齢	MMSE
		mean (SD)	mean (SD)
非認知症	12	74.5 (4.3)	27.5 (1.8)
認知症	12	75.9 (7.6)	21.2 (5.1)

認知症患者

- アルツハイマー病 (AD): 9
- 正常圧水頭症 (NPH): 1
- 軽度認知障害 (MCI): 1
- AD+NPH: 1

分類実験

- 特徴量の標準化
 - 平均: 0, 分散: 1
- 分類器
 - Linear SVM (SVM)
 - L1正則化付きLogistic Regression (LRL1)
 - L2正則化付きLogistic Regression (LRL2)
- モデル評価
 - Nested Leave-one-participant-out cross validation
 - ROC curve

自動検出の評価

	手動（書き起こし）		自動（音声認識）	
	AUC	Acc.	AUC	Acc.
MMSE	0.85	0.83	-	-
音声特徴量 (SVM)	0.85	0.83	0.74	0.75
音声特徴量 (LRL2)	0.90	0.83	0.85	0.79
音声特徴量 (LRL1)	0.97	0.96	0.88	0.71
音声 + 言語特徴量 (LRL1)	0.91	0.88	0.85	0.67
音声 + 言語特徴量 (LRL2)	0.92	0.92	0.71	0.58
音声 + 言語特徴量 (SVM)	0.95	0.92	0.71	0.58

自動検出手法の検出能力はMMSEと同等以上

自動検出のエラー分析

- 入力の特徴量・分類モデルに関係なく、
非認知症患者 2名， 認知症患者 2名を誤分類
(誤分類した認知症患者)
認知症の種類：AD, MMSEスコア：16, 25
- MCI (認知症の前段階) の方 (1名) は， 正しく分類

まとめ

目的

繰り返しの利用を想定した
認知症の自動検出

提案手法

質問をランダムで5問出題する
エージェントシステムによる自動検出

結果

MMSEと同等以上の検出能力

(提案手法) AUC: 0.88

(MMSE) AUC: 0.85

今後の課題

- 研究協力者数を増やして検証
- 認知症のタイプによる影響の確認
- 人対人, 人対エージェントの違いの確認

補足資料

早期発見の難しさ

本人



- 病識がない
- 受診拒否

周囲の人々



- 症状の見逃し
- 老化による症状との混同

定期的に状態を把握するツールが必要

診断基準 (DSM-V)

認知症

- 1つ以上の認知領域における機能の低下
(複雑性注意, 実行機能, 学習および記憶, 言語, 知覚-運動, 社会的認知)
- 認知領域の障害による**日常生活の阻害**



軽度認知障害 (MCI) (≡ 認知症の前段階)

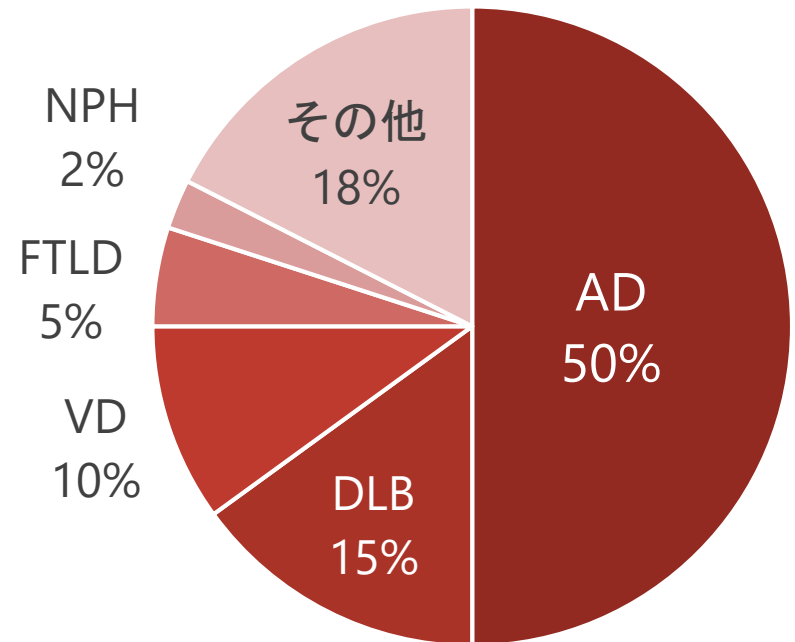
- 1つ以上の認知領域における機能の低下
(複雑性注意, 実行機能, 学習および記憶, 言語, 知覚-運動, 社会的認知)
- **日常生活の阻害がない**程度の認知領域の障害

認知症の病型

認知症は「症状」の集合

原因疾患

- アルツハイマー病 (AD)
- 正常圧水頭症 (NPH)
- レビー小体型認知症 (DLB)
- 前頭側頭葉変性症 (FTLD)
- 血管性認知症 (VD) など



認知症の特徴

記憶障害

- 最近の出来事の想起ができない（近時記憶の障害）
- 過去の記憶（遠隔記憶）は保たれていることが多い

見当識障害

- 時間や場所，人が分からなくなる

取り繕い反応

- 自身の能力低下を取り繕う

（例） Q. 今日は何曜日ですか？

A. この歳になると曜日は気にしないね.

中核症状とBPSD

中核症状：認知症の症状の主体をなす

記憶障害，見当識障害，失行，
失認，失語，実行機能障害など

性格

環境

BPSD（周辺症状）

幻覚・妄想，異常な食行動，
睡眠障害，易怒性，徘徊など

先行研究の課題

先行研究の 提案

(音声言語から認知症を検出するために)
有効な特徴量や検出手法

先行研究の 課題

課題が同じパターン
→ 繰り返しの利用に不向き

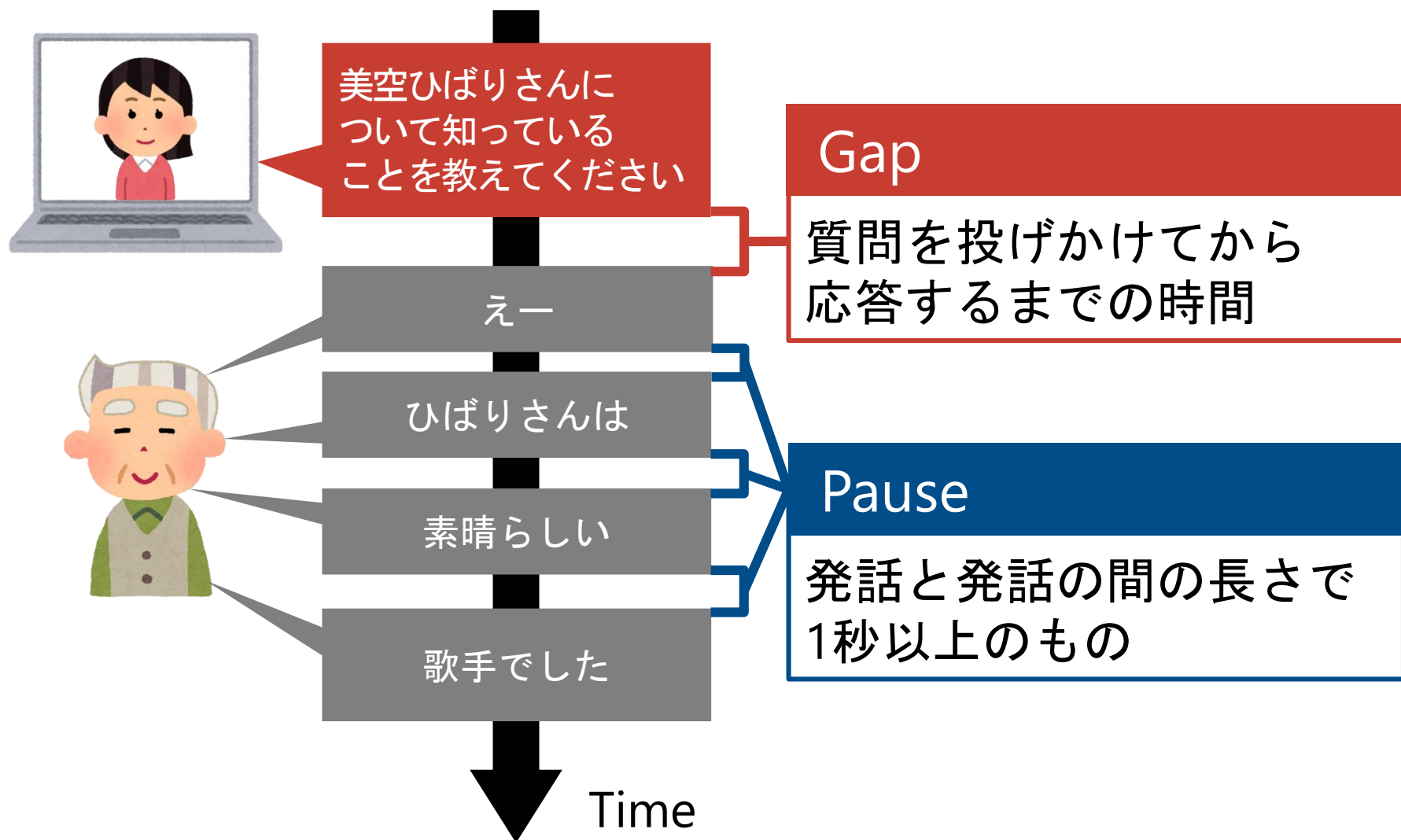
(音声言語から認知症を検出するために)
有効な質問に関する検討が不十分

繰り返しの利用を想定した
認知症の自動検出

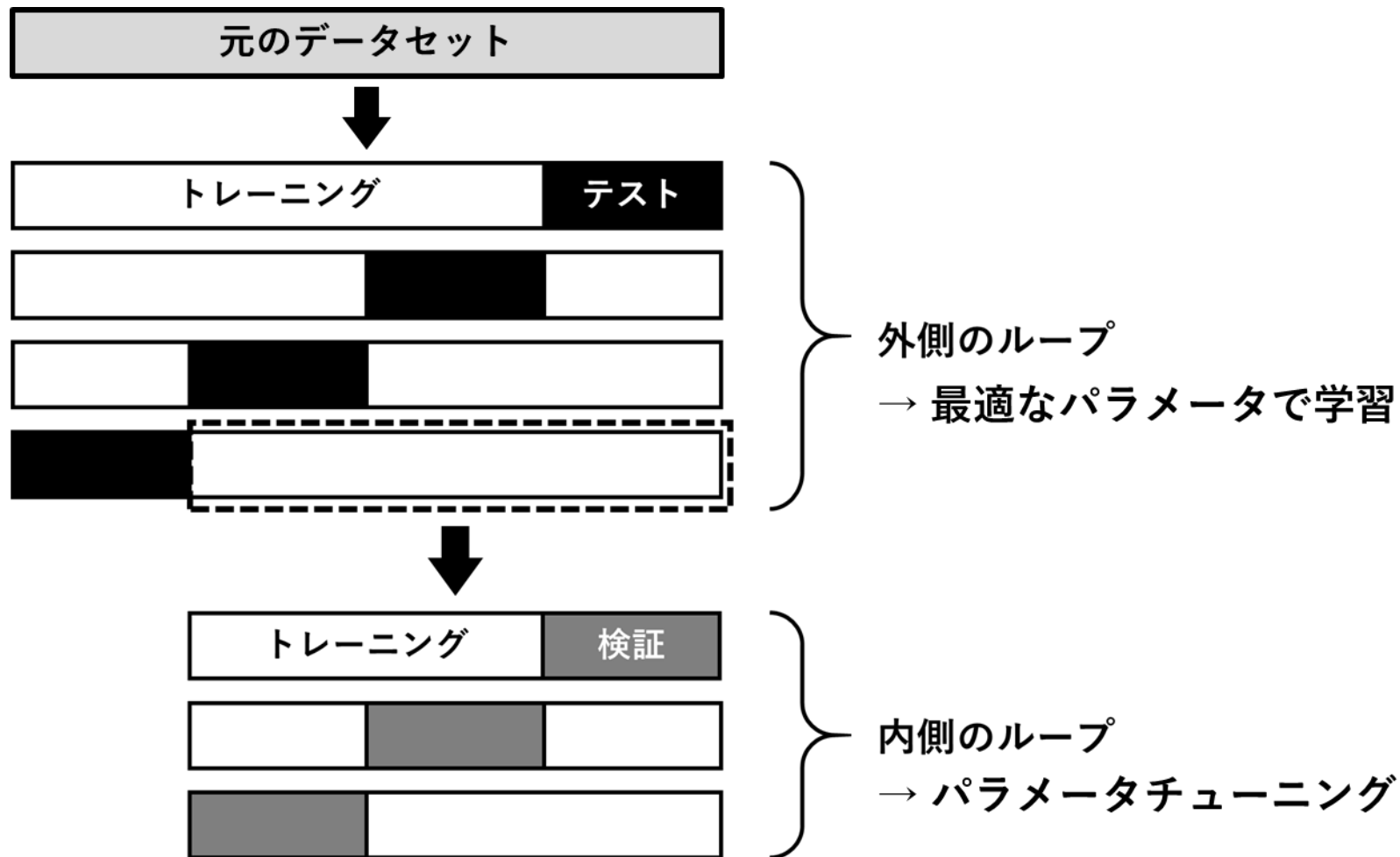
質問セット

Q1	ご家族のことを教えてください	人の見当識
Q2	日頃ストレスを感じていることを話してください	自己認識
Q3	趣味は何ですか	関心
Q4	好きな歌は何ですか	
Q5	石原裕次郎さんについて知っていることを教えてください	遠隔記憶
Q6	長嶋茂雄さんについて知っていることを教えてください	
Q7	美空ひばりさんについて知っていることを教えてください	
Q8	今の日本の総理大臣は誰ですか	神経心理学的検査
Q9	今の季節は何ですか	
Q10	今年は何年ですか	
Q11	あなたは左利きですか、右利きですか	自己認識
Q12	夜は眠れていますか	
Q13	食欲はありますか	

“Gap” と “Pause”



Nested cross validation



Gapの分析

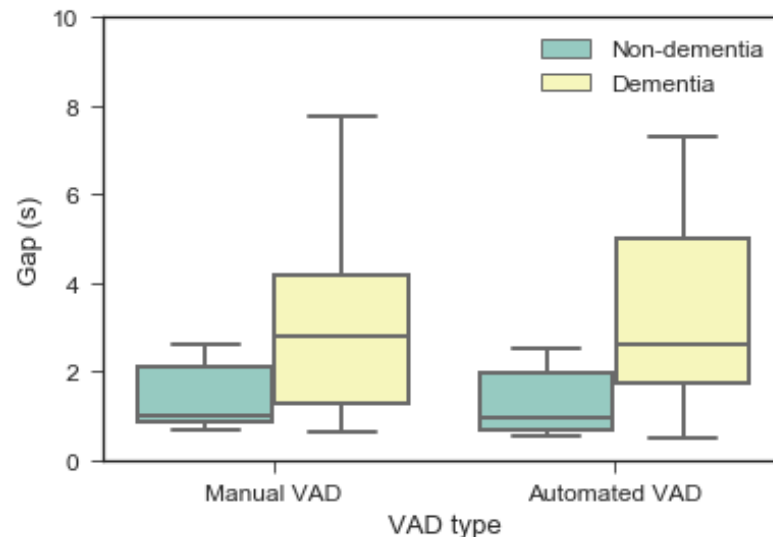
- 応答まで時間 (Gap)

非認知症者グループ < **認知症患者グループ**

- 2つのグループ間に有意差がみられた

Mann-Whitney's U test: $p < 0.05$ ($n = 24$)

Effect size: $r = 0.42$



VADの評価

- False acceptance rate (FAR)

$$FAR = \frac{N_{FA}}{N_{ns}}$$

N_{FA} : False acceptance frames, N_{ns} : Non-voice frames

→ 0.037 (3.7%)

- False rejection rate (FRR)

$$FRR = \frac{N_{FR}}{N_s}$$

N_{FR} : False rejection frames, N_s : Voice frames

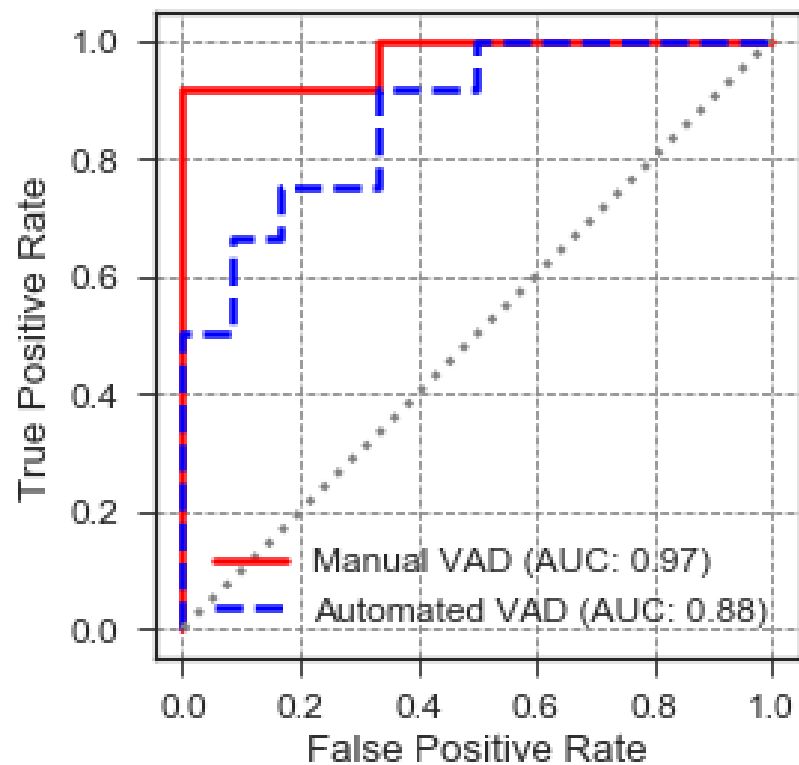
→ 0.073 (7.3%)

検出能力

		手動		自動	
		AUC	Acc.	AUC	Acc.
先行研究	MMSE	0.85	0.83	-	-
	音声+言語+画像特徴量 (SVM)	0.93	0.83	-	-
	音声特徴量 (SVM)	0.85	0.83	0.74	0.75
	音声特徴量 (LRL2)	0.90	0.83	0.85	0.79
	音声特徴量 (LRL1)	0.97	0.96	0.88	0.71
	音声 + 言語特徴量 (LRL1)	0.91	0.88	0.85	0.67
	音声 + 言語特徴量 (LRL2)	0.92	0.92	0.71	0.58
	音声 + 言語特徴量 (SVM)	0.95	0.92	0.71	0.58

提案手法の検出能力はMMSEと同等以上

ROC曲線



質問の変化による検出能力の変化

- 各モデルのうち、最高のAUC値を記載

質問数		1	2	3	4	5
手動	音声特徴量	0.59	0.81	0.76	0.99	0.98
	音声 + 言語特徴量	0.96	0.58	0.75	0.94	0.91
自動	音声特徴量	0.52	0.86	0.79	0.84	0.88
	音声 + 言語特徴量	0.56	0.90	0.90	0.90	0.85

LRL1の特徴選択による特徴量の分析

- すべての試行において選択された特徴量を抽出

音声＋言語特徴量		音声特徴量	
(手動)	(自動)	(手動)	(自動)
Gap	Gap	Gap	Gap
F0のレンジ	F0の中央値	F0のレンジ	F0の中央値
Pauseの平均値	Pauseの平均値	F0の中央値	
形容詞の数	形容詞の数	Powerの平均値	
発話速度		Powerの標準偏差	
Powerの平均値			
Powerの標準偏差			

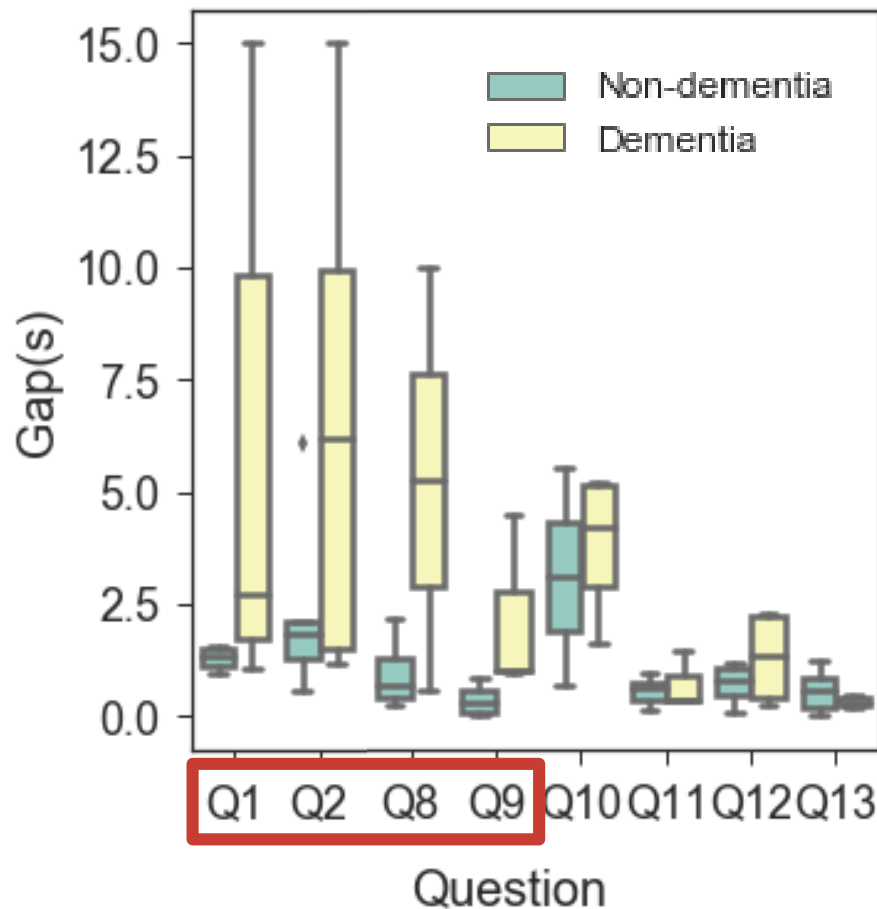
LRL2の重みによる特徴量の分析

TOP5 (音声+言語特徴量の分析)

	手動	自動
1	Gap	Powerの平均値
2	F0のレンジ	Gap
3	F0の最大値	F0の中央値
4	Pauseの平均値	Pauseの平均値
5	動詞の数	F0の標準偏差

Gapが有効な特徴量である可能性

質問によるGapの違い



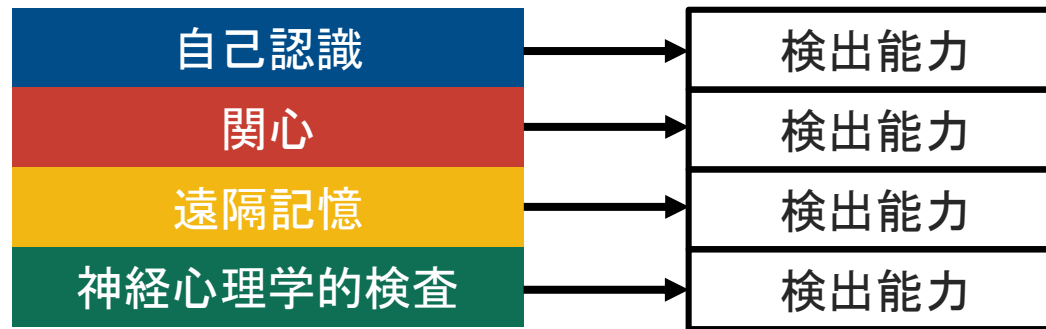
Q1	ご家族のことを教えてください
Q2	日頃ストレスを感じていることを話してください
Q8	今の日本の総理大臣は誰ですか
Q9	今の季節は何ですか

差がみられる質問

- 人の見当識に係る質問 (Q1)
- 自己認識に係る質問 (Q2)
(オープンエンド型質問)
- 常識に関する質問 (Q8)
- 時の見当識質問 (Q9)

質問の分析

- カテゴリ別に検出能力（AUC）を算出



※人の見当識に係る質問はデータ不足のため除外

- 各質問におけるGap（応答にかかった時間）をグループ別（認知症 / 非認知症）に算出

カテゴリごとの検出能力

カテゴリ	AUC	正解率
自己認識	0.67	0.62
関心	0.95	0.73
遠隔記憶	0.60	0.32
神経心理学的検査	0.63	0.73

- ・ 手動の書き起こしによる特徴量
- ・ 各モデルのうち最高のAUC値とその正解率を記載
- ・ 人の見当識に係る質問はデータ不足のため除外

- 関心に係る質問の検出能力が高い
- 遠隔記憶に係る質問の検出能力が低い

認知症患者の関心

- 早期段階から無関心や意欲低下が認められる
- アルツハイマー型認知症の典型的な症候



無関心や意欲低下は音声言語に反映されやすい



関心に係る質問において高い検出能力 (AUC: 0.95)

神経心理学的検査に基づく質問

- 一連の課題の総合的なスコアから判定
→ 一部分だけを抜粋しても、効果的とは限らない
- 回答の正誤がスコアに影響
→ 回答中の音声言語に着目して作成されていない



神経心理学的検査に基づく質問において
高くない検出能力 (AUC: 0.63)

ADと非認知症者の分類結果

AD, NPH, MCI
→ 出現しやすい症状が異なる



AD（9名），非認知症（12名）による分類

	手動（書き起こし）		自動（音声認識）	
	AUC	Acc.	AUC	Acc.
音声特徴量	0.95	0.86	0.89	0.67
音声＋言語特徴量	0.88	0.80	0.89	0.76

ADの検出能力も高い

Gapの分析

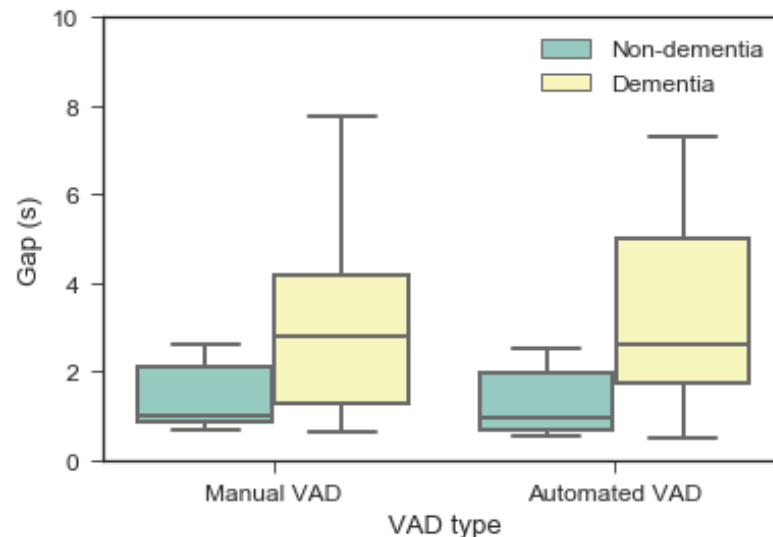
- 応答まで時間 (Gap)

非認知症者グループ < **認知症患者グループ**

- 2つのグループ間に有意差がみられた

Mann-Whitney's U test: $p < 0.05$ ($n = 24$)

Effect size: $r = 0.42$



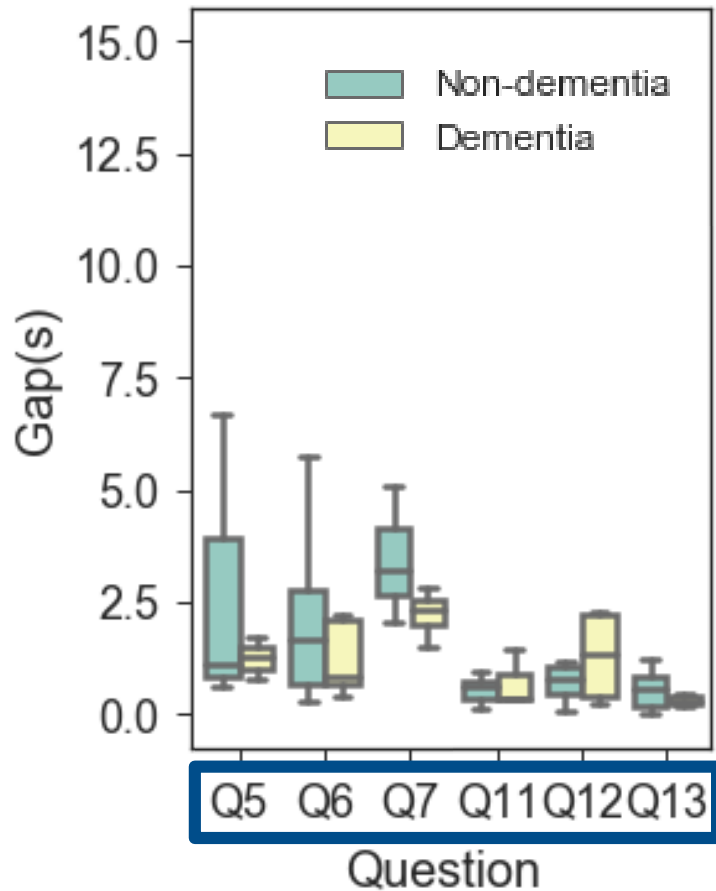
Gapの有効性

- Gapを抜いた場合の検出能力 (AUC) の変化

	手動	自動
音声特徴量	0.97	0.88
音声 + 言語特徴量	0.95	0.71
音声特徴量 (w/o Gap)	0.66	0.74
音声 + 言語特徴量 (w/o Gap)	0.67	0.81
Gapのみ	0.69	0.58

Gapは有効な特徴量の一つ
Gapのみで検出することは難しい

質問によるGapの違い



Q5	石原裕次郎さんについて知っていることを教えてください
Q6	長嶋茂雄さんについて知っていることを教えてください
Q7	美空ひばりさんについて知っていることを教えてください

遠隔記憶に係る質問



低い検出能力

遠隔記憶に係る質問

早期段階では遠隔記憶は保たれていることが多い



研究協力者の遠隔記憶が障害されていない

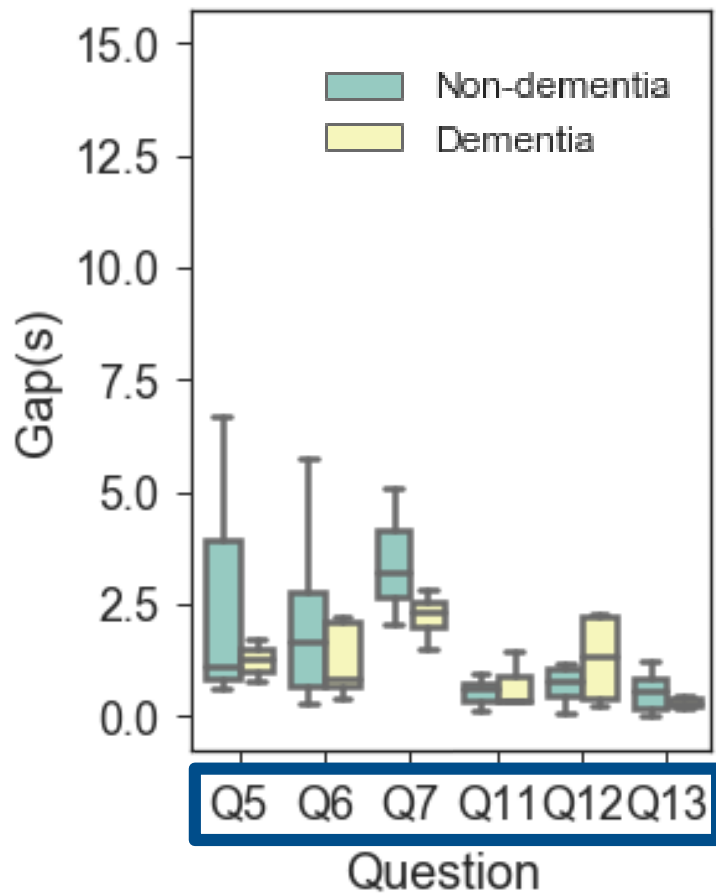


音声言語に反映されない



遠隔記憶に係る質問において低い検出能力 (AUC: 0.60)

質問によるGapの違い



Q11 あなたは右利きですか，左利きですか

Q12 夜は眠れていますか

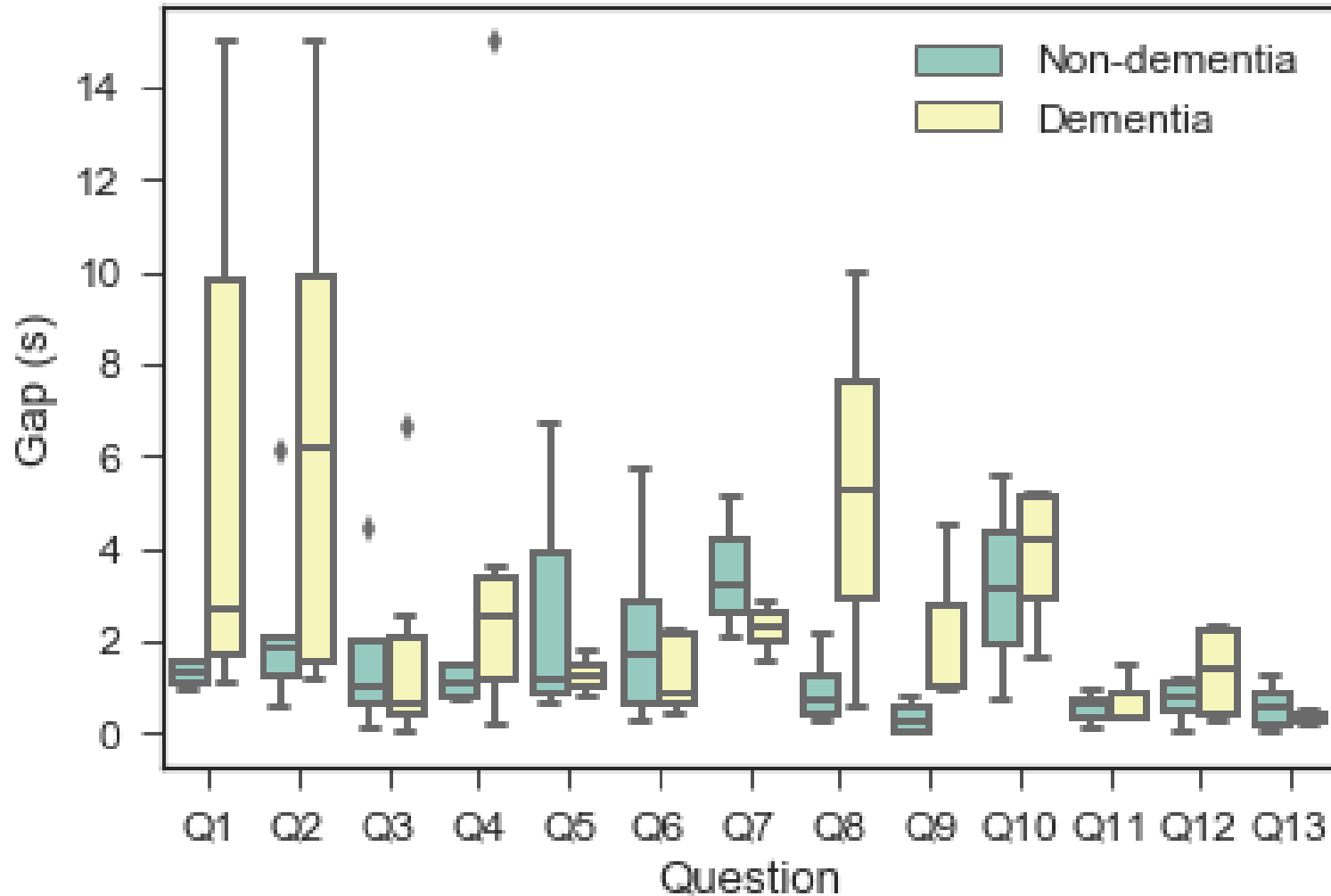
Q13 食欲はありますか

自己認識に係る質問
(クローズド型質問)

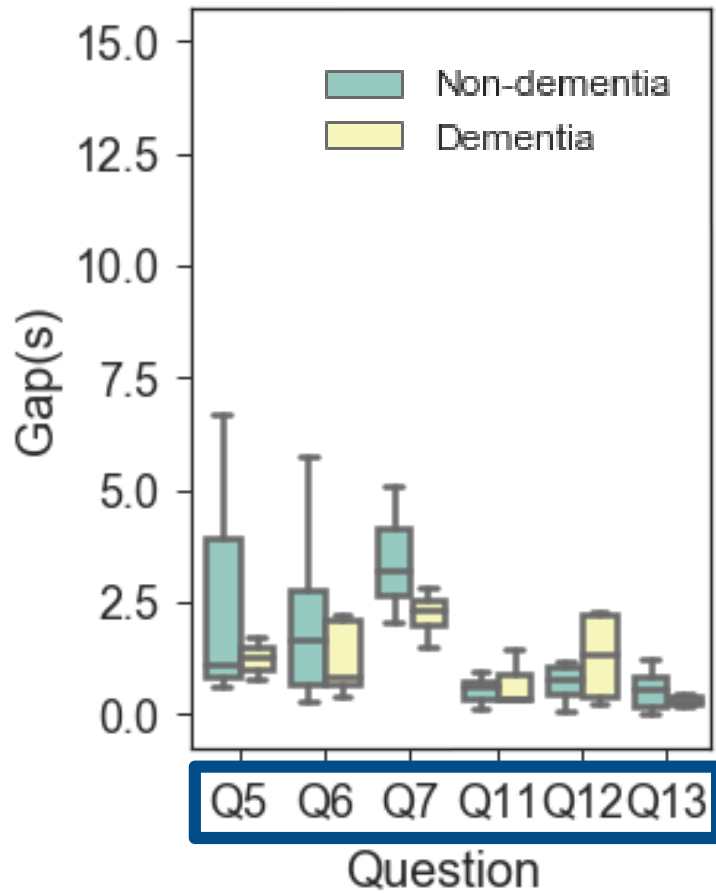
回答が容易

Gapに差がみられない

質問によるGapの違い



質問によるGapの違い



Q5, Q6, Q7, Q11, Q12, Q13には
差がみられない