

令和 2 年 6 月 26 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K00251

研究課題名(和文) テンソル空間を利用したパターン特徴抽出・識別方式の研究

研究課題名(英文) Pattern extraction and recognition using tensor space

研究代表者

新田 恒雄 (Nitta, Tsuneo)

早稲田大学・グリーン・コンピューティング・システム研究機構・その他(招聘研究員)

研究者番号：70314101

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：機械学習の進展はビッグデータを持つ様々な情報システムを実現した。一方、脳波などデータ蓄積が少ない分野では、少ない量のデータから本質的内部構造を解析可能な、新しい計算モデルが望まれている。本研究では適度なデータ量で対象の内部構造を、(A)対象クラス内のベクトル構造表現(多重線形写像)、及び(B)複数クラス間のベクトル構造表現(クラス間外積空間)から解析する手法を調査・検討した。同時に、解析手法を発話時脳波信号の分析・特徴抽出・音節/単語識別に適用し、手法の有効性を確認した。一連の研究成果から、最終目標である「音声想起時脳波からの音声言語情報抽出・識別」研究を推進する手掛かりを得ることができた。

研究成果の学術的意義や社会的意義

比較的少ないデータ量で解析対象の内部構造や複数クラス間の関係を抽出する技術は、ビッグデータに基づく機械学習の恩恵を受けられない分野では重要になると考えられる。ここで研究された内容は、今後、脳-コンピュータインターフェース(BCI)を開発する際の基礎技術となると考えている。音声想起BCIが実現すると、考えた内容をスマートフォンに転送することが可能になるため、将来、ALS患者の方達が情報発信可能な端末を実現できる。

研究成果の概要(英文)：Machine learning technology has been realized many AI systems that have a big and open data set. On the other hand, there exists many systems that could not utilize accumulated database like brain-computer interfaces (BCIs) and these areas require a new computing-model that enables us to extract internal structures and/or analyze deep structures with a moderate amount of data. In this study, the multiple linear mapping (MLM) technology is applied to automatic speech recognition, and the wedge product technology is applied to brain (EEG) signals when uttering.

研究分野：知能情報学

キーワード：テンソル積 多重線形写像 脳波解析 外積空間 DNN

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

(1) Deep Neural Network (DNN)など機械学習技術の進展により、ビッグデータを前提とする情報システムの研究・開発の進展が著しい。一方、脳波のようにデータ蓄積の少ない分野では、適度な量のデータから本質的内部構造を解析可能な、新しい計算モデルの開発が望まれる。

(2) 合成写像に基づくベクトル-ベクトル変換を実現する方法として、圏論(category theory)と呼ばれる新しい数学分野が発展しつつある。一方、圏論はデータ処理全体の鳥瞰図を与えるが、写像を引き起こす具体的な機能(関手 functor)の設計は、個々の応用分野に依存している。

2. 研究の目的

(1) 適度なデータ量から対象の本質的な内部構造を:

(A) 対象クラス内のベクトル構造表現から、および

(B) 複数クラス間のベクトル構造表現から

解析可能な手法を調査・検討する。

(2) 上記の解析手法を脳波解析に応用し、将来の「音声想起時脳波からの音声言語情報抽出・識別」研究を推進する手掛かりを探る。

3. 研究の方法

(1) 脳波(*Electroencephalogram*; EEG) 信号を中心に以下の(2)と(3)の技術項目について、それらの有効性を検証するため(一部、音声信号による検証を含む)、必要な前処理技術:

(1-1)ノイズ除去技術,

(1-2)ノイズに強い分析技術,

を開発する。

(2) クラス内のベクトル構造表現に関しては、次の二つの技術を脳波解析に適用する。

(2-1) KL 変換 (KL transform; KLT)

(2-2) 多重線形写像(multiple linear mapping; MLM テンソル積)

(3) クラス間のベクトル構造表現に関しては、クラス間外積空間を脳波解析に適用する。

4. 研究成果

(1) 脳(EEG)信号に関する前処理技術の開発

報告者も長年関与した音声認識研究開発と比較すると、脳信号からの音声言語識別は研究が始まって間もない分野である。そこで、上記した技術を比較検討するに当たっては、ノイズ除去方式、分析方式などの前処理技術の開発が必要となった。ここでは以下の二つの技術成果を説明する。

(1-1) 脳波中のノイズおよび artifact 除去

脳波(EEG)データは、微小信号のため電源ノイズや α 波などの低周波帯域信号との分離がSN比向上に不可欠である。一般的に電源ノイズ除去では、50Hz or 60Hz の notch filter が用いられることが多い。しかし電源ノイズは高調波を含むため、単一周波数のフィルタでは除去しきれない。我々は、EEG 信号のうち 70Hz 以上の high γ 帯を対象に分析を行うが、種々の定常雑音を除去するため、予め音声言語と無関係な区間(発話や想起に遡る区間)で EEG 信号スペクトラム $N(p, k, t)$; p - 電極番号, k - 周波数, t - 時間フレーム, を観測し、認識対象 $X(p, k, t)$ から引去る noise-spectrum subtraction を導入した。

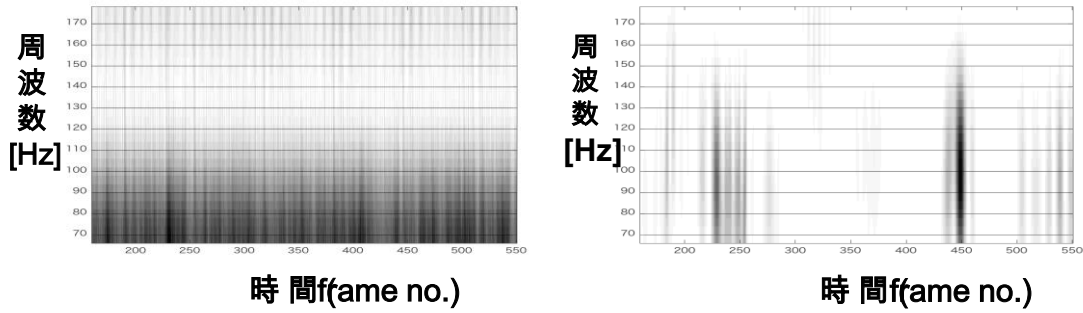


図 1 ノイズスペクトラム除去前 (左) 及び 除去後 (右)

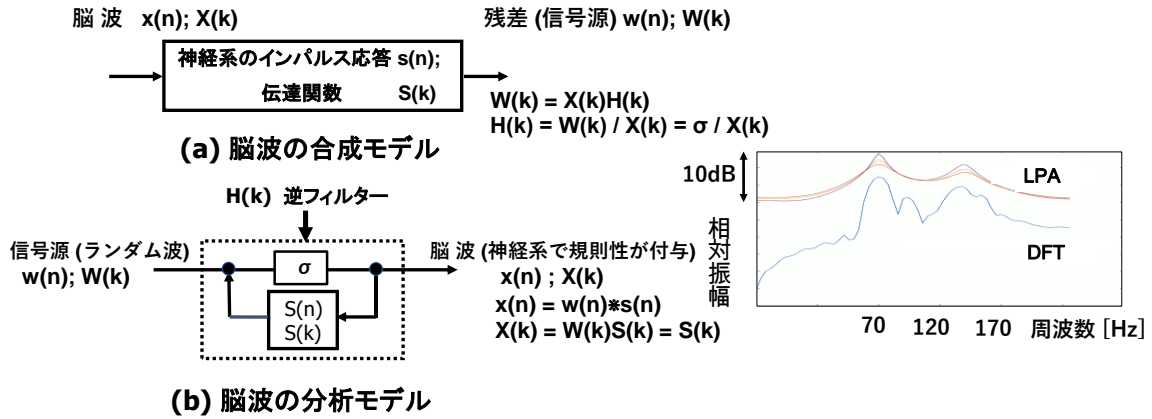


図 2 線形予測分析 LPA (左) 及び LPA と FFT スペクトラムの比較 (右)

$$X(p, k, t) - N(p, k, t) \rightarrow X(p, k, t) \quad (1)$$

図 1 に、離散フーリエ変換(Discrete Fourier Transform; DFT) 後、電極毎にノイズスペクトルを引去った時の例を示す。なお、周波数帯域は 70Hz ~ 180Hz の帯域を使用している。これにより主に低周波数帯域に観測される、筋肉動作を中心とする artifact の影響を避けることができた。この後、逆 DFT (IDFT)により各電極信号は波形に戻され、発話時の脳波(60ch.) については、60 個の実効値(root mean square; rms)が計算され、空間分布として使用される。また、音声想起(発話なし)時の脳波は、次の LPA による分析が行われる。

(1-2) 線形予測分析 (linear predictive analysis; LPA)

脳波を対象とした LPA モデルを図 2 に示した。合成モデルでは、信号源としてのランダム波入力 $w(n)$ と、言語情報を生成する神経系のインパルス応答 $s(n)$ との畳み込み演算から、言語の規則性が付与された脳波 $x(n)$ が出力される。図には周波数領域での表現 $\{W(k), S(k), X(k)\}$ も示した。インパルス応答 $s(n)$ は Levinson-Darvin 線形予測係数 $\{a_0, a_1, \dots, a_p\}$ から計算することができる。電極で観測される脳波 $x(n)$ に対する分析モデルでは、合成モデルで与えられた線形予測係数から、脳波の持つ言語表象を計算する (Analysis-by-Synthesis; AbS)。通常の計算アルゴリズムでは、まず脳波の自己相関係数から Levinson-Durbin 法を適用して線形予測係数を求め、逆フィルタを構成して脳波の LPA スペクトルを計算する。この時、線形予測係数(ここでは 8 次)に 0-padding を施すことで、任意の周波数分解精度を得ることができる。図 2 の右に /a/ を想起した際の分析結果を DFT と比較して示した。

(2) クラス内のベクトル構造を表現する

(1)で開発した脳波信号に対する前処理技術を活用し、次の二つのクラス内ベクトル射影技術を検討する。

(2-1) KL 変換 (KLT)の検討

KLTは定常雑音などに対する直交展開関数として提案されたが、DFTの周波数成分間に直交関係がない場合(音声では声道の共鳴周波数が現れるため、DFTあるいはその圧縮版のBPF(帯域通過フィルタ)群は成分間の直交性に欠ける)にもKLTが利用される。またDCT(離散コサイン変換)は計算が簡単なため、KLTの近似として利用されることが多い。KLTは主成分分析(PCA)の固有ベクトルを基底として展開するため、スペクトル時系列の二次元パターンに対しても適用することができる。その場合、周波数軸と時間軸に沿う一次微分及び二次微分パターンを観測することができる。

発音時の脳波では、運動野を中心に構音(or 調音)の違いが空間パターンの違いとして現れることが分かっている(K. E. Bouchard et al., Nature, 2013)。本研究では、発音時の脳波(60 ch.)のrmsパターンからKLT成分の時系列を得ることで、18単音節の認識に50%を超える性能を得ることができた(神崎ほか, 日本音響学会春季講演会, 2017)。この時の音節分類器は部分空間法(Subspace method; SM)を用いている(KLT/SM)。

(2-2) 多重線形写像(multiple linear mapping; MLM or テンソル積)の検討

同一クラスのスペクトル時系列内の構造をMLMにより陽に抽出することができる。図3は、双線形写像(bi-linear map)により、テンソル成分をDNN(Deep Neural Network)の入力に利用した例を示している。DNNは最初にAuto-encoderとして圧縮された特徴を抽出し、続いてスタックに積まれた特徴から音素を分類してHMM(隠れマルコフモデル)に送る。ここでは、英語連続音声の大量データを使用して、評価実験を行った結果を図4に示した(T. Ogawa 他, Interspeech2015)。実験は、認識評価に使用されることのできる多いTIMITコーパスを使用した。図から、従来の音声特徴(BPF, MFCC)に双線形写像(bilinear map)を施すことで誤認識を大きく改善できること、また複数の特徴を併用することでさらに性能が改善し、その場合にも

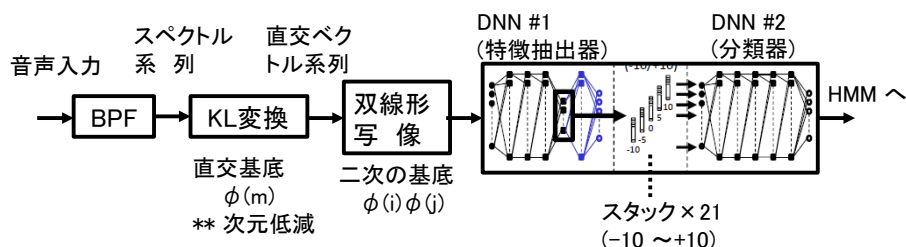


図3 bi-linear (テンソル積)を組み込んだDNN/HMMシステム

テンソル空間を利用する写像は、更に大きな改善を示すことが分かる。この結果は、それまで知られていたDNN/HMMの最良評価値に匹敵している。

(3) クラス間のベクトル構造を表現する

(2-1)で説明したKLT/SMは、誤認識が同じ子音グループ(Ca-Ci-Cu-Ce-Co, 子音Cは母音グループ#を含む)の母音間で生じていた。SMはクラス内の固有ベクトルから類似度を計算する。そこで、母音の違いを表現するため、異なるクラスの二つのベクトル間の相違性を外積(楔積(wedge product; WP))で表現することを検討した。具体的には母音(例えばba)毎に、他の母音(bi, bu, be, bo)との間で楔積を計算する。ここでbaは学習デー

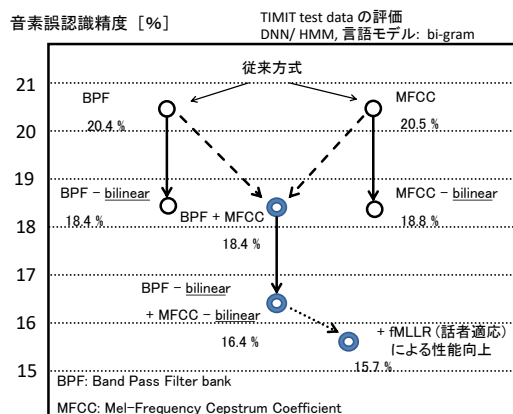


図4 英語音声認識でのbi-linear適用効果

タ, 対する(誤認識し易い) bi, bu, be, bo は代表ベクトルとして, 子音グループ毎に次の楔積を求めた.

$$\text{b-group: } \{ Y^{ba} \wedge Y^{bo}, Y^{ba} \wedge Y^{bu}, Y^{ba} \wedge Y^{bi}, Y^{ba} \wedge Y^{be} \},$$

$$\text{d-group: } \{ Y^{da} \wedge Y^{do}, \text{-----}, Y^{da} \wedge Y^{de} \},$$

$$\text{g-group: } \{ Y^{ga} \wedge Y^{go}, Y^{ga} \wedge Y^{gu}, Y^{ga} \wedge Y^{gi}, Y^{ga} \wedge Y^{ge} \},$$

$$\text{v-group: } \{ Y^a \wedge Y^o, Y^a \wedge Y^u, Y^a \wedge Y^i, Y^a \wedge Y^e \}$$

この計算では, 最初に group 毎に楔積(右三角行列; 各成分は行列中の行列式として表現される)を4種類計算する. 学習データと代表ベクトルとの楔積は, テンソル成分として互いに浸み出した成分が評価され, この成分を双線形空間として固有ベクトルを4種類求める.

次に認識段階では, まず Frobenius norm から, 最も小さい音節を対立するものとして選択し(同じベクトルであれば, norm はゼロになる), 続いて, 未知データと選択された代表ベクトルから楔積を計算して, 部分空間法と同じく固有ベクトルとの類似度を計算して評価する(WP/SM; 図5参照). KLT/SM(50.4%)単独からさらに WP/SM を加えてハイブリッド方式とすることで, 性能が向上する(56.1%). なお, 二方式の類似度 S1, S2 は, hybrid similarity として重み加算で $\alpha S1 + (1-\alpha)S2$ で評価している.

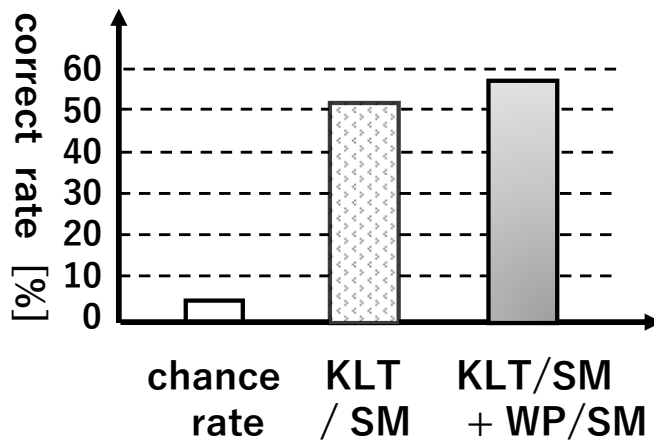


図5 各手法間の性能比較 [%]

以上, 主に発話時脳波 (EEG) 信号を中心に, 比較的少量のデータで高い性能を達成可能なクラス内ベクトル構造表現, 及びクラス間のベクトル構造表現を検討・評価した結果を述べた. 今後, 最終目標である音声想起時脳波からの音声言語情報識別を目指したい.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 入部百合絵, 北岡教英	4. 巻 Vol.73 No.5
2. 論文標題 音声認識に向けた超高齢者音声のコーパス構築	5. 発行年 2017年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 303-310
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Satoshi Tamura, Hiroshi Ninomiya, Norihide Kitaoka, Shin Osuga, Yurie Iribe, Kazuya Takeda, Satoru Hayamizu	4. 巻 Vol. E99-D No.10
2. 論文標題 Investigation of DNN-based audio-visual speech recognition	5. 発行年 2016年
3. 雑誌名 IEICE Transactions on Information and System	6. 最初と最後の頁 2444-2451
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Seng Kheang, Kouichi Katsurada, Yurie Iribe and Tsuneo Nitta	4. 巻 Vol. E99-D, No.4
2. 論文標題 Using Reversed Sequences and Grapheme Generation Rules to Extend the Feasibility of a Phoneme Transition Network-based Grapheme-to-Phoneme Conversion	5. 発行年 2016年
3. 雑誌名 IEICE Transactions on Information and System	6. 最初と最後の頁 1182-1192
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計29件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Meiko Fukuda, Ryota Nishimura, Norihide Kitaoka, Hiromitsu Nishizaki, Yurie Iribe
2. 発表標題 Construction of a Corpus for Elderly Japanese Speech Recognition
3. 学会等名 IEEE GCCE 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Norihide Kitaoka, Yurie Iribe, Hiromitsu Nishizaki
2. 発表標題 Construction of a Corpus of Elderly Japanese Speech for Analysis and Recognition
3. 学会等名 LREC2018 (国際学会)
4. 発表年 2018年

1. 発表者名 梶並 和明, 西村 良太, 入部 百合絵, 北岡 教英
2. 発表標題 音声対話破綻検出手法の開発に向けた音声対話データ収録システム
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 秋水紫苑, 入部百合絵, 北岡教英
2. 発表標題 対話システムにおける非言語情報を用いた対話破綻の検出
3. 学会等名 情報処理学会 第81回全国大会
4. 発表年 2019年

1. 発表者名 曾我真子, 入部百合絵, 兒嶋朋貴, 増田達明
2. 発表標題 個体識別を目的とした牛音声の音響情報分析
3. 学会等名 情報処理学会 第81回全国大会
4. 発表年 2019年

1. 発表者名 梅澤舞菜, 入部百合絵, 北岡教英
2. 発表標題 方言を考慮した音声言語情報に基づく高齢者認知症傾向の検出
3. 学会等名 情報処理学会 第81回全国大会
4. 発表年 2019年

1. 発表者名 梅村直人, 入部百合絵
2. 発表標題 発音動作可視化を目的とした口腔形状の特徴量抽出
3. 学会等名 情報処理学会 第81回全国大会
4. 発表年 2019年

1. 発表者名 平田里佳, 深井健太郎, 入部百合絵, 桂田浩一, 新田恒雄
2. 発表標題 発話時脳波のスペクトル成分を用いた音声言語情報の識別
3. 学会等名 第16回情報科学ワークショップ
4. 発表年 2018年

1. 発表者名 小口 優人, 大村 英史, 桂田 浩一
2. 発表標題 Active Appearance Modelsを用いた読唇
3. 学会等名 第5回サイレント音声認識ワークショップ
4. 発表年 2018年

1. 発表者名 深井 健太郎, 大村 英史, 桂田 浩一, 平田 里佳, 入部 百合絵, 新田 恒雄
2. 発表標題 発話時脳波を利用した音声言語情報の識別
3. 学会等名 第5回サイレント音声認識ワークショップ
4. 発表年 2018年

1. 発表者名 森口寛生, 大村英史, 桂田浩一
2. 発表標題 変分オートエンコーダーを用いた多重音解析の性能評価
3. 学会等名 情報処理学会第80回全国大会
4. 発表年 2018年

1. 発表者名 新田恒雄, 桂田浩一, 神崎卓丸
2. 発表標題 脳波による日本語短音節の認識
3. 学会等名 第4回サイレント音声認識ワークショップ
4. 発表年 2018年

1. 発表者名 平田里佳, 入部百合絵, 新田恒雄
2. 発表標題 英語発音学習に向けた音素連結・脱落同化パターンの分析と検出
3. 学会等名 日本音響学会2018春季研究発表会講演論文集
4. 発表年 2018年

1. 発表者名 川島愛美, 入部百合絵, 北岡教英
2. 発表標題 高齢者の対話音声から抽出した言語的・音響的特徴に基づく認知症傾向の判別
3. 学会等名 電子情報通信学会東海支部卒業研究発表会
4. 発表年 2018年

1. 発表者名 平田里佳, 入部百合絵, 新田恒雄
2. 発表標題 英語発音学習に向けた音素連結・脱落同化パターンの分析と検出
3. 学会等名 教育システム情報学会東海支部学生研究発表会
4. 発表年 2018年

1. 発表者名 神崎卓丸, 浅原康平, 中根丈司, 桂田浩一, 杉本俊二, 堀川順生, 新田恒雄
2. 発表標題 発話時と想起時の脳波による日本語短音節認識の比較
3. 学会等名 日本音響学会
4. 発表年 2017年

1. 発表者名 浅原康平, 中根丈司, 神崎卓丸, 桂田浩一, 杉本俊二, 新田恒雄, 堀川順生
2. 発表標題 日本語音節発話・想起時の脳波解析
3. 学会等名 日本音響学会
4. 発表年 2017年

1. 発表者名 黒川有紀, 入部百合絵, 北岡教英
2. 発表標題 音響的特徴を利用した高齢者の認知症傾向の分析
3. 学会等名 日本音響学会
4. 発表年 2017年

1. 発表者名 澤田優希, 入部百合絵, 北岡教英
2. 発表標題 マルチモーダル情報を用いた運転中におけるシステム向け発話の推定
3. 学会等名 日本音響学会
4. 発表年 2017年

1. 発表者名 桂田浩一
2. 発表標題 Suffix Arrayを用いた高速STDにおけるキーワード分割の最適化に関する検討
3. 学会等名 日本音響学会
4. 発表年 2017年

1. 発表者名 Kohei Asahara, Jozi Nakane, Takumaru Kanzaki, Shunji Sugimoto, Kouich Katsurada, Tsuneo Nitta, and Junsei Horikawa
2. 発表標題 EEG during Japanese syllable recall and speech tasks
3. 学会等名 Proc. of The 3rd Annual Meeting of the Society for Bioacoustics
4. 発表年 2016年

1. 発表者名 Takumaru Kanzaki, Shunji Sugimoto, Kouich Katsurada, Junsei Horikawa, and Tsuneo Nitta
2. 発表標題 Japanese monosyllable recognition from EEG
3. 学会等名 Proc. of The 3rd Annual Meeting of the Society for Bioacoustics
4. 発表年 2016年

1. 発表者名 浅原 康平, 中根 丈司, 神崎 卓丸, 中澤 香太, 桂田 浩一, 杉本 俊二, 新田 恒雄, 堀川 順生
2. 発表標題 日本語単音節発話時と想起時の脳波解析
3. 学会等名 日本音響学会
4. 発表年 2016年

1. 発表者名 神崎 卓丸, 浅原 康平, 中根 丈司, 中澤 香太, 桂田 浩一, 杉本 俊二, 堀川 順生, 新田 恒雄
2. 発表標題 脳波からの日本語単音節認識方式の検討
3. 学会等名 日本音響学会
4. 発表年 2016年

1. 発表者名 Yurie Iribe, Norihide Kitaoka, Shuhei Segawa
2. 発表標題 Speech corpus spoken by young-old, old-old and oldest-old Japanese
3. 学会等名 Proc of. LREC2016
4. 発表年 2016年

1. 発表者名 深井 健太郎, 大村 英史, 桂田 浩一, 平田 里佳, 入部 百合絵, 新田 恒雄
2. 発表標題 音声想起時脳波からの音節情報の抽出
3. 学会等名 人工知能学会
4. 発表年 2020年

1. 発表者名 田口 亮, 付 明川, 新田 恒雄
2. 発表標題 音声想起脳波データラベリング効率化のための半自動音節アライメント
3. 学会等名 人工知能学会
4. 発表年 2020年

1. 発表者名 山尾 元陽, 平田 里佳, 入部 百合絵, 深井 健太郎, 桂田 浩一, 新田 恒雄
2. 発表標題 脳波信号を用いた音声想起時単語の認識
3. 学会等名 人工知能学会
4. 発表年 2020年

1. 発表者名 深井 健太郎, 大村 英史, 桂田 浩一, 平田 里佳, 入部 百合絵, 付 明川, 田口 亮, 新田 恒雄
2. 発表標題 音声想起脳波からの言語表象抽出と音節認識
3. 学会等名 電子情報通信学会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	入部 百合絵 (Iribe Yurie) (40397500)	愛知県立大学・情報科学部・准教授 (23901)	
研究 分担者	桂田 浩一 (Katsurada Kouichi) (80324490)	東京理科大学・理工学部情報科学科・准教授 (32660)	