公開

研 究 終 了 報 告 書

「脳からの言語情報解読技術の開発」

研究期間： 2018 年 10 月～2022 年 3 月
研 究 者： 堀川 友慈

1． 研究のねらい

本研究はヒトの脳活動から詳細な言語情報を解読する技術を開発し、脳を介した柔軟なコミュニケーションを実現するための基盤技術の確立を目指すものである。そのために脳と人工知能技術を統合した脳からの言語情報デコーディング技術の開発を行う。

脳情報デコーディングは、脳活動から心や身体の状態を解読する技術であり、脳とコンピュータを直接つなぎ身体を介さない機械の操作や意思伝達を可能とする「ブレイン・マシン・インターフェイス（brain-machine interface, BMI）」の基盤となっている。この技術は、身体を介さない意思伝達手段を提供可能であり、身体の制約を受けないより自由なコミュニケーションを可能とする未来の情報通信技術として期待が持たれている。しかし、従来の機械学習を用いた手法では、学習データを取得するための実験的制約に起因する、多様かつ柔軟な出力を生成することが困難であるという問題がある。

そこで本研究では、研究代表者が行ってきた深層ニューラルネットワークを脳から活用する技術を発展させ、脳を介した情報伝達性能を向上させることで、脳情報通信技術をより柔軟に利用できる社会システム基盤の構築に貢献することを目指す。そのために、脳と人工知能技術を統合したデコーディング技術を言語情報の解読に応用し、より詳細な脳の情報をより多様な刺激条件下におけるfMRI脳計測信号から解読する技術を開発することを目標とする。 具体的には、以下の3点の実現に取り組む。

1．解読情報の詳細化：脳活動に表現されている情報を単語としてだけでなく文章として解読することを可能にする。

2．解読モダリティの多様化：画像提示中の脳活動だけでなく、音声やテキスト刺激提示中の脳活動など多様な条件下の脳活動から知覚や心的内容の脳情報解読を可能にする。

3．脳における言語表現の神経基盤の解明：脳からの言語情報デコーディング技術の開発を通して、画像や音声などの感覚刺激が概念的意味表現に変換される過程に関する神経基盤を明らかにする。

本研究開発は、脳情報の詳細な解読を可能にし、脳を介した情報伝達の性能を向上させることで、BMI 技術を必要とする人々により柔軟な情報伝達手段を提供する。これにより、身体的制約によって周囲とのコミュニケーションが困難な人々であっても、身体的困難を感じずに適応可能な社会の構築に貢献することが期待される。

2． 研究成果

（1）概要

本研究では、脳から詳細な言語情報を解読する技術を開発することを目的とし、機能的磁気共鳴画像法（functional magnetic resonance imaging, fMRI）による脳計測データから、刺激のモダリティによらず被験者の知覚内容を文として解読するための技術開発を行った。この目的を

達成するため、脳活動から深層ニューラルネットワーク（deep neural network, DNN）への信号変換技術を応用し、DNN を介した脳からの言語生成アプローチの提案とその有効性の検証を行った。具体的には、近年、自然言語処理分野で開発が進められている深層言語モデル（deep language model, DLM）の特徴量を用いて、さまざまな刺激の知覚中や想像中の脳の意味情報に関する情報表現の解析を行った。さらに、脳から DLM の特徴量を予測し、予測特徴量を文生成モデルによって処理することで、脳に表現されている知覚・心的内容を言語化する技術の開発を行った（図1）。本研究ではまず、DLM の特徴量がヒトの脳で表現される意味情報を捉えるために有効であるかを検証するため、動画観察中の脳活動を刺激由来の特徴量（動画キャプションから抽出した意味特徴）から予測する脳活動予測解析を用いて、代表的な DLM である BERT モデルの特徴量と視覚物体や単語の意味を表現するモデルによって計算した特徴量との間で、脳活動予測成績の比較を行った。その結果、脳の広範な部位の活動を DLM が他のモデルよりも高い精度で説明できることがわかった。この結果を受けて、提示動画のキャプション文に対して DLM の複数の階層から計算した特徴量を、動画観察中の脳活動から予測し、得られた予測特徴から文の生成が可能かどうか検証を行った。その結果、脳から生成した文が、提示動画につけられたキャプション文と高い類似性を示すことが確認できた。さらに、動画刺激提示中の脳活動だけでなく、音声刺激やテキスト刺激知覚中の脳活動や、動画を想像中の脳活動に対する脳活動予測、脳活動からの特徴量予測解析を行うことで、刺激のモダリティによらず DLM の特徴量が脳の意味情報表現と高い類似性を示すことを確認することができた。以上の結果は、脳に表現されている詳細な意味の情報を、DLM の特徴表現を活用することで捉えることが可能であることを示唆しており、脳が表現する意味情報を捉えることで、思考内容を直接言語化する脳情報通信技術の可能性を開拓するものであるといえる。

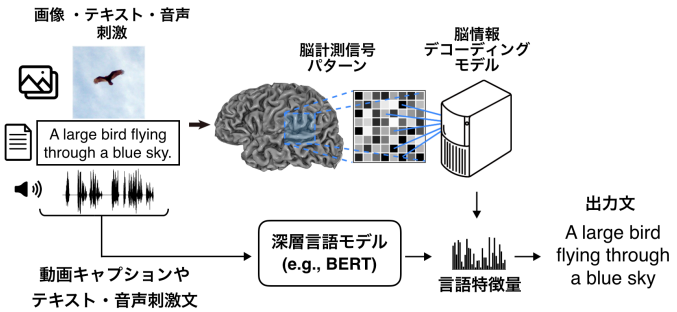

図 1. 脳からの言語情報解読アプローチの概要。脳から予測した深層言語モデルの言語特徴量をもとに、脳に表現されている意味情報の言語化を行う。

（2）詳細

本研究では、「脳からの言語情報解読技術の開発」という目標のもと、1.解読情報の詳細化、2.解読モダリティの多様化、3.脳における言語表現の神経基盤の解明、の 3 つの研究項目を実施した。以下では、説明のため研究項目 3,1,2 の順で概説する。

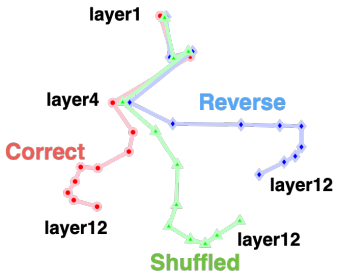

図 2. 情報表現類似度解析。多次元尺度法による類似度の二次元で可視化したところ高次の階層が単語順の入れ替え（shuffled, reverse）に大きく影響を受けた。

研究項目 3:「脳における言語表現の神経基盤の解明」

本研究項目では、知覚対象に関わる言語情報および詳細な意味情報に関するヒト脳の神経基盤を解明するため、深層言語モデル(deep language model, DLM)の特徴量を用いた脳活動解析を行った。この解析に先立ち、まず提示刺激中のどのような情報が、脳の各領域の情報表現と関連しているかを調べるため、動画刺激に対して、動画中の物体・風景・イベントなどの意味情報に関連するラベル（dog, building, explosion など）と、動画刺激に対する観察者の感情のラベル（joy, fear など）をタグ付けし、動画の各フレームを入力とした DNN 特徴、意味ラベル、感情ラベルのそれぞれから、脳活動の予測モデルの構築および予測解析を行ない、各脳部位（ボクセル）がどのモデルから最も高い精度で予測できるかを評価した。その結果、後頭部に位置する視覚野から、脳の前方に向かって、視覚モデル、意味モデル、感情モデルの順に緩やかに分布が変化している様子を確認することができた。この結果は、先行研究によって視覚刺激由来野の脳活動をよく説明することが知られていた物体カテゴリの情報よりも、ここで使用した意味的な情報、さらには感情的な情報が、より多感覚に関連する脳部位で表現される情報であることを示唆するものである（Horikawa et al., 2020）。この結果を受け、高次の意味情報を扱うモデルとして DLM に着目し、DLM の特徴量がヒトの脳で表現される意味情報を捉えるために有効であるかを検証するため、動画観察中の脳活動を刺激由来の特徴量（動画キャプションからの言語特徴抽出）から予測する脳活動予測解析を用いて、代表的な DLM である BERT と視覚物体や単語の意味を表現するモデルによって

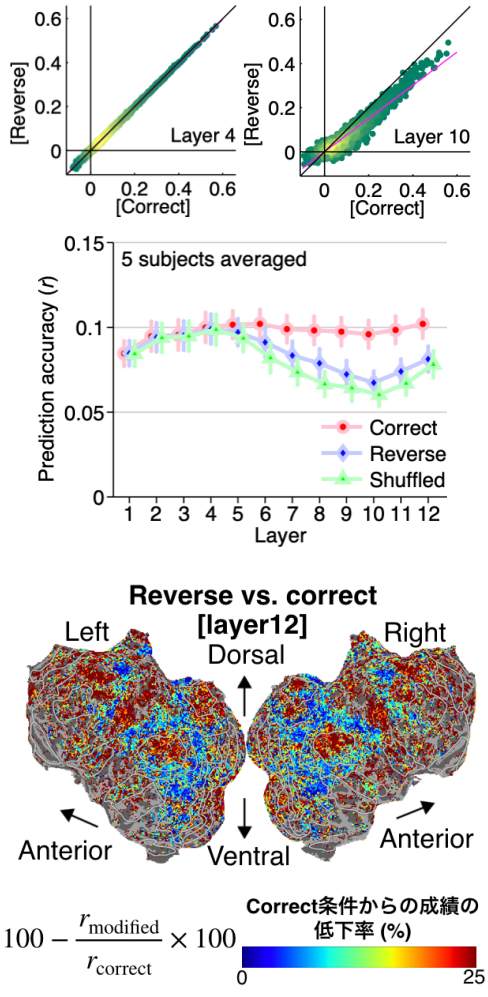

図 3. BERT への入力文の単語順を変化させた時の脳活動予測成績への影響。（上）正しい単語順の文に基づく成績と逆順の単語順の文に基づく予測成績の比較例。（中）全ボクセルの平均予測成績の比較。（下）予測成績の低下率の皮質マップ。

計算される特徴量による脳活動予測成績の比較を行った。その結果、脳の広範な部位の活動をDLMが他のモデルよりも高い精度で説明できることがわかった。さらに、DLM の特徴量が文脈的な情報を表現しているかを確認するため、モデルへの入力文の単語順を入れ替えて文脈情報を失わせる操作が、モデルの特徴量や脳の予測成績にどのような影響を与えるかを調べたところ、DLMの高次層が文脈情報に対してより感度が高く(図 2)、これらの階層からの脳活動予測において、角回(angular gyrus)などをはじめとする default mode network と呼ばれる脳の高次の領域の活動の予測成績が影響を受けることを確認できた(図 3)。この結果は、DLM の高次層が脳における文脈を考慮した意味情報の表現の有効なモデルとなることを示すとともに、脳の高次の領域において知覚要素間の関係性などの文脈情報が表現されていることを示唆している。

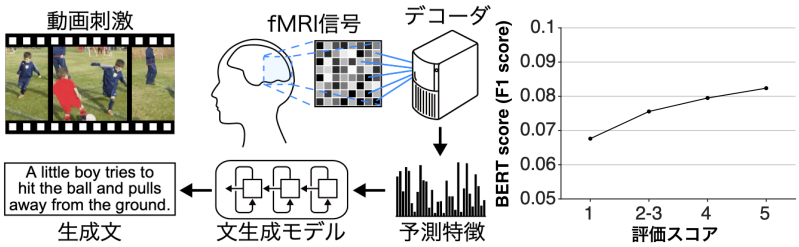

図4. 動画観察中の脳活動から文生成。(左)脳からの文生成の概要。脳活動から予測した深層言語モデルの特徴量に基づく文生成を行った。(右)被験者自身によるキャプション文の評価スコアごとの脳からの生成文と候補キャプション文の類似度スコア。評価には BERT score の F1score を用いた。

研究項目 1:「解読情報の詳細化」

本研究項目では、脳からの解読情報をより詳細化する形で脳情報解読技術の効用を向上させるため、深層言語モデルの特徴を介したデコーディング解析を行った。従来研究においては、視覚物体単体や個々の単語の意味情報の解読は実現されているが、知覚対象の意味の要素をなす単語間の関係性や交互作用などを表現可能な文レベルの詳細な情報の解読は行われていなかったため、このような文脈情報を考慮した意味情報を解読可能かどうかの検討を行った。上記の DLM の特徴を用いた脳活動予測解析によって、DLM の高次層の特徴が単語間の順序関係に基づく文脈情報を表現しているということが明らかになったことを受け、脳から解読した特徴量においても、このような文脈依存の意味情報が表現されているかの確認を行った。具体的には、脳のさまざまな部位から予測した DLM の特徴量のパターンを、正しい単語順の文に基づく特徴パターンとランダムに単語順を入れ替えた文に基づく特徴パターンと比較し、前者を同定できるかを調べる同定解析を行った。その結果、比較的高次の特徴量を解読したときに、高い同定成績が得られるという、上記の他の解析と一貫した結果が得られた。さらに、DLM の言語特徴から文を生成する文生成モデルを構築し、動画観察中の脳活動から予測した特徴量から動画内容を説明する文へ変換する解析を行った結果(図 4 左)、脳から生成した文が、複数の評価者によって提示動画につけられたキャプション文と高い類似性を示すことが確認できた。さらに脳からの生成文が被験者自身の主観的体験を反映しているかを確認するため、被験者自身に、自身が見ていた動画内容と複数の他の評価者によって作成されたキャプション文の整合性を評価させたところ、脳から解読した特徴量や生成文が、被験者が高い評価を与えたキャプション文と高い類似性を示すことがわかった(図 4 右)。これらの結果は、単なる単語の集合としてではなく、文レベルの意味の情報を脳活動から解読することが可能であ

ることを実証するとともに、脳に表現されている意味情報を脳活動から直接言語化（文として生成）することが可能であることを示唆するものである。また、これまでの視覚階層における脳活動が主観的な情報を反映しているという知見（Horikawa & Kamitani, in press）を、意味情報のレベルにまで拡張し、脳の情報表現が被験者の主観体験と強く結びついていることをさらに強く支持する結果であるといえる。

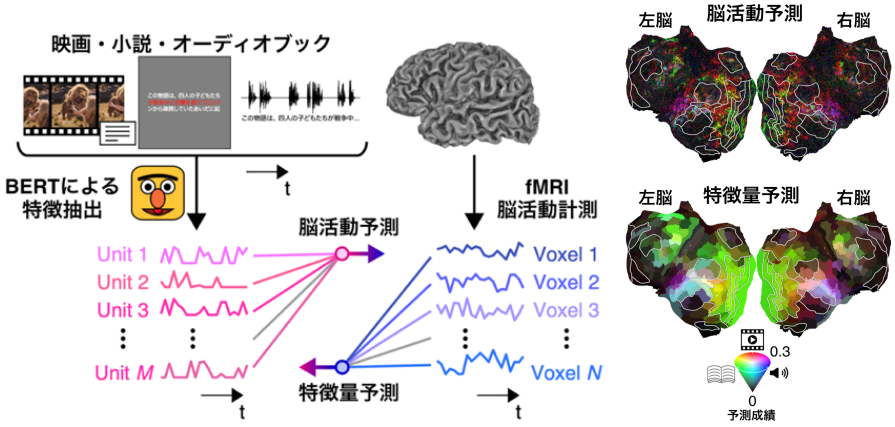

図5. 自然な言語関連刺激を用いた脳活動予測および言語特徴量予測解析。（左）異なる刺激モダリティに対する深層言語モデル特徴を使った脳活動予測と特徴量予測解析の概要。（右）脳活動予測および特徴量予測結果の皮質マップ。

研究項目 2:「解読モダリティの多様化」

本研究項目では、脳からの意味情報の解読を特定の刺激モダリティに限定されず、画像・動画や音声・テキスト刺激など、多様な刺激モダリティに対して適用できるよう、解読対象となる刺激モダリティを多様化することが可能かどうかを検討した。この目的のため、同一の作品に対する映画、小説、オーディオブックを刺激として被験者に提示している時の脳活動を計測し、それぞれ動画の２秒ごとのクリップにつけたキャプション文やテキスト・音声として提示した文から DLM（BERT）を用いて抽出した言語特徴と、刺激提示中の脳活動との間で、脳活動予測および特徴量予測解析を行った（図5左）。その結果、脳活動予測においても特徴量予測解析においても一貫して、左右の上側頭回や、前頭前野の一部など、刺激の意味的な情報表現に関して、モダリティ間で共通性があるという従来の知見と整合性のある部位から比較的高い成績の結果が得られた。しかしながら、特定のモダリティ刺激条件で訓練したモデルを用いて他のモダリティ条件下での脳活動を解析する汎化解析を行ったところ、刺激のカバーする範囲の狭さが問題となり、さほど高い成績は得られなかった。そこで、解析対象を視覚的な映像に対する描写である動画のキャプションに絞り、刺激を実際に知覚しているときと、キャプション文に基づいて視覚映像を想像している時の脳活動間でモデルの汎化解析を行った結果、特に深層言語モデルの特徴を使った際に、知覚と想像間で高い汎化性能が得られ、脳活動からの特徴量予測解析においては，知覚と遜色のない想像内容の意味情報の解読が可能であることが確認できた。この結果は、適切に解析対象となる刺激空間を合わせることで、刺激の有無に関わらず、脳内で表現されている詳細な意味情報を捉えることができることを示唆している。

以上の結果は、脳に表現されている詳細な意味の情報を、DLM の特徴表現を活用することで捉えることが可能であることを示唆するとともに、本研究を通して、DLM 特徴を介して脳が表現する意味情報を解読することで、思考内容を直接言語化する脳情報通信技術の可能性を開拓することができたといえる。

3. 今後の展開

本研究では、深層ニューラルネットワークと脳情報解析を統合する技術を言語情報の解読に応用することにより、従来解析手法以上により詳細な言語・意味に関わる脳情報の解読や、そのような意味情報表現の神経基盤の理解を深めることができた。これにより、脳からの言語情報解読技術を様々な特性を有する人々の脳活動解析に適用することで、個々人の脳の特性をより詳細に調べる研究を展開することが可能になる。具体例を挙げると、近年、「知覚機能は正常だが想像することが困難である」というアファンタジアという特質を持つ人々の存在が注目されつつあり、行動実験や質問紙を通して、アファンタジアの人々が視覚や聴覚などの感覚に関わる想像を苦手とする一方で、想像能力の欠如を言語能力により補っているという知見が報告されている。本研究で開発した言語情報の脳情報解読技術は、研究代表者がこれまでに開発してきた視覚情報の脳情報解読技術と組み合わせることで、このような人々の脳の情報処理特性をより網羅的に調べることに活用することができる。また、本研究において、脳からの解読情報が、被験者の主観を反映していることを示唆する結果が得られたが、この結果は、本研究で開発した技術を、例えば芸術的な素養を持つ人と素人が芸術作品を鑑賞する際の目の付け所の違いや認識の深さの違いを脳活動から定量化することや、個々人の言語能力の違いが脳における知覚情報処理にどのような違いを生じさせるかという古典的な問いに最新の技術を用いて切り込むことを可能にする。このような方向性の研究は、今後数年以内に実現可能となると期待される。一方で、本研究によって、知覚対象や想像対象の意味の情報が、脳の広範な部位によって表現されていることが明らかになった。したがって、本研究で開発した技術を日常的に利用可能な社会システムに組み込むためには、脳全体の活動を簡易に計測する技術の開発が必要不可欠であり、このような技術の社会システムへの実装は今後十年単位での技術開発が必要となると考えられる。脳と深層ニューラルネットワークなどの AI 技術の融合という観点では、ここ数年の視覚や聴覚の感覚情報処理から言語情報処理への技術の発展の速さを鑑みると、今後数年でより高次の認知情報処理である、記憶や感情の計算モデルや脳情報解析への応用が進展してくことが見込まれる。

4. 自己評価

本研究課題の遂行により、研究開始当初掲げていた 3 つの研究項目の検討を通して、深層言語モデルの特徴表現を活用することで、単語の集合としての意味情報を超えて、文による記述レベルの意味情報を、脳活動から捉えることが可能であることを確認することができた。脳からの文生成および多様なモダリティからの言語情報解読という課題に関しては、いまだその精度の面で十分な結果が得られているとは言い難いが、脳から解読した情報が、被験者自身の主観的体験を反映した情報であることや、刺激を提示しない想像課題中の脳活動からも、知覚条件と遜色のない精度で言語特徴の解読が可能であることを確認することができた。また、より具体的な課題として、脳から解読した言語・意味情報を実用的なコミ

ュニケーションに活用するためには、脳からの生成文の文として自然さを向上させることが必要となることや、受動的な動画・音声・テキスト知覚が、必ずしも高い精度での脳情報解読を可能にするだけの共通の意味情報表現を惹起しないという問題点の確認をすることができた。社会システムをデザインするという観点からは、fMRI を用いた脳計測という大掛かりな実験を必要とする性質上、即座に脳情報通信の実用化への応用や、社会や経済に波及効果をもたらす技術開発には至らなかったが、本研究開発により、従来の脳解析手法で得られていた以上に高い精度で、脳に表現される知覚・想像内容の意味・言語情報を定量化することが可能になった。上述の通り、この技術は、近年注目を浴びている「知覚能力は正常だが想像することが困難である」という特質を有するアファンタジアと呼ばれる人々の脳の特性を調べるために活用したり、個人間の言語能力の違いや注目する対象の違いなどに基づく脳の情報表現の違いを定量化したりするためのツールとして有効に活用することが期待される。この点において、本研究開発は、さきがけの事業の「独創的・挑戦的かつ国際的に高水準の発展が見込まれる先駆的な基礎研究を推進し、社会・経済の変革をもたらす科学技術イノベーションの源泉となる、新たな科学知識に基づく創造的な革新的技術のシーズ（新技術シーズ）を世界に先駆けて創出することを目的とするネットワーク型研究(個人型)」という趣旨を遵守することができたと考えられる。研究の進め方に関しては、最終年度に所属機関の移動や、コロナによる実験実施の制約などが重なり、予定通りの研究体制や研究費執行とはならなかったが、研究費を使って購入した物品や実施した実験および収集したデータは適切かつ最大限に有効活用することができたといえる。

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：2件

1. Horikawa, T., Cowen, A.S., Keltner, D., & Kamitani, Y. ″The neural representation of visually evoked emotion is high-dimensional, categorical, and distributed across transmodal brain regions″ iScience 23, 101060. doi: https://doi.org/10.1016/j.isci.2020.101060 (2020)
ヒトの脳内で視覚、意味、感情の情報がどのように表現されているか調べるため、動画刺激の視覚、意味、感情に関連する特徴量を用いた脳活動解析を行った。Principal gradient と呼ばれる脳全体の階層構造の指標と関連づけて評価したところ、単一感覚モダリティと複数感覚モダリティの違いを強く反映している gradient 1 の軸に沿って、視覚、意味、感情の順によく予測できる脳部位の分布のピークがシフトしていることが明らかになり、意味の情報を解読するために有効な特徴や解析対象とすべき脳部位を明らかにすることができた。
2. Horikawa, T. & Kamitani, Y. ″Attention modulates neural representation to render reconstructions according to subjective appearance″ Commun. Biol. 5, 34 https://doi.org/10.1038/s42003-021-02975-5 (2022)
刺激由来のボトムアップの視覚特徴の情報に対して、注意によるトップダウンの情報がどのように階層的な視覚特徴に影響を与えるのかを調べた。脳からの DNN 特徴の予測解析と予測特徴からの画像再構成解析を通して、DNN の階層構造に沿って、視覚野の各階層の情報

表現が特異的な注意の影響を受けることや、脳からの解読情報が被験者の主観的な見えを反映していることを示した。この結果は、脳で表現される情報が被験者本人の主観を反映していることを示唆するものであり、意味情報の解読において解読結果が被験者の主観を反映していることを視覚特徴のレベルで補完するものである。

（2）特許出願

研究期間全出願件数：0件（特許公開前のものも含む）

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

1. 堀川友慈. “機械の脳で読み解くヒトの心の神経基盤” 第 29 回脳の世紀シンポジウム, オンライン （September, 2021）
2. Horikawa, T. ″Neural representations of visually evoked emotional experience″ The symposium of 'Revealing relationships among cognitive functions, among emotions, and among concepts using neuroscience-based approaches', Japan (December, 2020).
3. Horikawa, T. ″Neural mechanisms underlying mental imagery: Brain decoding of mental images during imagery, dreaming, and attention″ The 21st annual meeting of the Japanese Imagery Association, Japan (November, 2020)
4. Horikawa, T. ″Brain decoding of mental contents via deep neural network features″ The 43th annual meeting of the Japan Neuroscience Society, Japan (July, 2020)
5. Horikawa, T., & Kamitani, Y. ″Attention biases neural representations of hierarchical visual features″ CCN2019, Berlin, German (September, 2019)
6. Horikawa, T. ″Decoding of What’s Not There from Human Brain Activity″ What’s Not There. A Workshop of Hallucinations, Dreams, Imaginations, and Virtual Reality., Toronto, Canada (June 2019)