

2019年7月10日

(報道発表資料)

日本電信電話株式会社

音を認識するために訓練された深層ニューラルネットワークが脳における音の表現と類似した表現を獲得することを発見

～動物の脳が進化の過程で音認識に適した性質を獲得してきた可能性を示す～

日本電信電話株式会社(本社:東京都千代田区、代表取締役社長:澤田純、以下 NTT)は、音認識のために訓練された深層ニューラルネットワーク(DNN)が哺乳類の脳と類似した音の表現を獲得することを発見しました。

神経生理学者の長年の研究により、音に対する神経細胞の反応特性が、様々な動物種の脳において明らかとなってきています。本研究では、こうして観察されてきた神経細胞の性質を、なぜ脳が持つようにいったのか、という疑問に答えます。本研究ではDNNを自然音認識で訓練し、それを構成する素子の性質を調べる際に、動物の神経細胞を対象とした神経生理実験と同じ方法で分析する技術を用いました。その結果、脳と類似した音の表現が獲得されていることを発見しました。

本研究の発見から、脳が進化の過程で音の認識処理に適応した神経表現を獲得したことが示唆されます。また、本技術により、脳研究と人工知能研究のさらなる融合が期待されます。

掲載雑誌:Journal of Neuroscience

掲載時期:7月10日(米国東海岸標準時)予定(確認中)。

【研究の背景と経緯】

ヒトを含む動物(哺乳類)の脳では、音が耳に届いてから認識されるまでに、脳幹から大脳皮質にいたる多段階の処理で音の特徴が分析されます。音の特徴の中でも、振幅変調(振幅の緩やかな変化)は音の認識に重要な手がかりです。これまでの長年に渡る神経生理学の研究は、神経細胞が「どのように」振幅変調を表現しているのか、という疑問の答えを、聴覚神経系における多数の脳部位について明らかにしてきました。しかしながら、神経細胞が「なぜ」そのように振幅変調を表現するようになったのか(必然性はあるのか)、という疑問については、神経細胞の性質と進化の過程との関係を一般的な実験的アプローチでは確かめることが原理的に難しく、それに答えることができませんでした。

実験的なアプローチの問題を補完するものとして、計算機モデルによるシミュレーションは有効なアプローチかもしれません。ただし、従来の典型的な計算機モデルによるアプローチでは、音の特定の特徴に対して特定の神経回路の性質を詳細に再現することはできても、聴覚神経系の重要な機能である自然な音の認識との関係に言及することはできません。

【研究の内容】

近年、人工ニューラルネットワークを用いることで、自然で複雑な音の認識を実現できるようになってきました。本研究は、この技術を導入することによって前述の疑問にアプローチしました(図1)。人工ニューラルネットワークのなかでも、特に深層ニューラルネットワーク(DNN)は多数の素子からなる多数の層が縦列した構造をしている点で、聴覚神経系と同様な構造を持っているといえます。しかし、その点以外は、聴覚系の具体的な神経回路を模擬したものではありません。

もし、このDNNが、動物と同様に自然な音を認識するよう訓練された結果として動物の脳と類似した特徴を獲得するならば、それは逆に、動物の脳が持つ性質が進化の過程で音認識に適応した結果である可能性を示すものといえるでしょう。

本研究では、自然音の分類課題のためにDNNを訓練したうえで、そのDNNを動物の脳を調べる神経生理学実験を模した方法によって分析しました。具体的には、訓練したDNNに様々な変調周波数(変動の繰り返しの速さ)を持つ音を入力し、DNNを構成する個々の素子について出力を調べました(図1)。その結果、素子によっては特定の変調周波数にのみ強く反応することや、その反応特性が処理段階が進むとともに規則的に変化することなど、動物の聴覚神経系について過去の研究で報告されているものと同様な性質がDNNに発現していることがわかりました。また、脳との類似性がDNNの訓練が進む過程で徐々に獲得されること、音の認識精度が高いDNNほど類似性が高いこと、自然音の認識に訓練しなかったDNNでは類似性が見られないこともわかりました。

以上の結果は、動物の脳において観察されてきた振幅変調の表現も、音の認識に適した性質が進化してきた過程で獲得されたものであるという可能性を示唆するものと言えます。

【今後の展開】

音の認識に用いられる音の特徴は、振幅変調以外にも多くあります。今後は、人工知能研究の知見をさらに取り入れながら、振幅変調以外の多くの特徴についても、人工ニューラルネットワークと脳とを比較し、動物の脳が進化してきた過程についてより一般的な知見を得ることを目指します。

【技術のポイント】

(1) 過去の神経生理学の研究結果のメタ分析

振幅変調は、ヒトを含む動物が音を認識するための重要な手がかりです(図 2)。このため、これまでに多くの神経生理学の研究によって神経細胞による振幅変調の表現が調べられてきました。

典型的な細胞では、神経反応としてのスパイク発火(図1左)が、振幅変調波形に同期して生ずることがよく知られています。スパイク発火の時間的パターンが振幅変調の情報を表していることから、この表現形式は「時間符号」と呼ばれます。スパイク発火が同期できる変調周波数には上限があります。このような変調周波数を上限周波数と呼びます。また、細胞によっては特定の変調周波数にのみ強く同期するものもあります。このような性質を変調チューニングと呼び、細胞の同期性を最大にする変調周波数を最適変調周波数と呼びます。一般に、神経細胞の属する脳部位が末梢から中枢になるにしたがって、上限周波数や最適変調周波数が低下する傾向があります。

変調波形への同期の有無にかかわらず、神経細胞の平均的な反応強度(スパイクの発火率)が変調周波数とともに変化することもあります。この発火率による変調情報の表現形式は「発火率符号」と呼ばれます。変調波形の情報をそのまま伝達しているわけではない点で、発火率符号は時間符号よりも抽象的な表現といえます。時間符号の場合と同様に、神経が発火する変調周波数には上限があったり、変調チューニングがみられることもあります。しかし、このような特性が見られる神経細胞は、一定の処理段階以上の部位にしか存在しません。

私たちは、過去の神経活動の研究結果をメタ分析することにより、この傾向を可視化すると同時に、DNN における振幅変調の表現との定量比較を可能にしました(図 3)。

(2) 聴覚神経系のモデルとしての人工ニューラルネットワーク

ヒトを含む動物の聴覚系は、多数の神経細胞からなる多数の脳部位が段階的に縦列している構造を持っており、音の認識もこの聴覚神経系によって実現されます。本研究では、人工ニューラルネットワークの一種である深層ニューラルネットワーク(DNN)が聴覚神経系と同様に、多数の素子からなる多数の層が縦列した構造によって音を認識することに着目し、DNNによって聴覚神経系をモデリングできるのではないかと考えました。

音声認識などの実用的な目的で用いられている人工ニューラルネットワークでは、認識精度を最大化するために音波形を前処理して計算した特徴量が入力としてよく用いられます。学術目的の聴覚モデル研究においても、内耳の出力を模擬した信号を取り出すといった前処理を行ったうえで、それをモデルへの入力とすることが典型的なアプローチです。しかし、こういった前処理の中身や精度は、研究者が事前に設定する仮説によって決まり、このために結果の解釈に影響を与える恐れがあります。

本研究では、前処理を適用しない音を直接入力として受け取り、一般的な音の分類課題を行う DNN を構築しました。そうすることで、研究者が事前に設定する仮説を最大限に省きながら、聴覚神経系の末梢から中枢までの全ての段階を統一的にシミュレーションできるようになりました。本研究では、各層で信号の時間分解能を損なわない、dilated convolutional network と呼ばれるタイプの DNN を用いました。

このDNNを、自然な音の分類課題のために訓練しました。自然音の分類課題は、どの動物種にとっても生存に重要であると考えられるため、動物の進化の過程との関係を一般的に示すのに適しています。訓練により、このDNNは、自然な音をある程度良い精度で分類できるようになりました。

(3) DNNを対象とした神経生理学実験

訓練したDNNにおける振幅変調の表現を、過去の神経生理学の研究結果と直接比較するために、DNNに神経生理学の手法を適用して振幅変調の表現を分析しました(図4)。神経生理学の手法を模し、DNNに正弦波状に振幅変調した音を入力し、その時の各素子の活動から入力音の振幅変調への同期性と平均活動強度を計算しました(それぞれ時間符号と発火率符号に対応)。その結果、(1)で述べたような聴覚神経系に見られる特性と類似の傾向が、DNNに発現することがわかりました。具体的には、変調チューニングを示す素子が存在すること、発火率符号によるチューニングが上位の層において出現すること、などがあげられます。さらに、入力層から出力層にかけて素子の最適変調周波数や上限周波数が系統的に変化する点でも類似することが確認されました(図4)。聴覚神経系の部位と、DNNの層の間の類似度を定量化したところ、DNNの縦列全体が、末梢から中枢までの聴覚神経系の全体と類似していることがわかりました(図5)。このように、末梢の聴神経から中枢の大脳皮質までの全経路に渡って、DNNにおける音の表現との対応関係を明らかにしたのは、世界で初めてです。

また、音の分類精度が高いDNNほど、脳との類似性が高い傾向もわかりました。これは、聴覚系の神経細胞が持つ性質が、音の情報処理の点で合理的であることを示唆するものです。

本研究は、機械学習に関する技術的な観点からも意義があります。DNNをはじめとする人工ニューラルネットワークが、高精度でデータ(音や画像など)を認識できることは様々な分野で実証されてきました。しかし、その認識がどのようなデータ表現に基づいて行われているのかは、いまだに難しい問題です。これまでに、人工ニューラルネットワークにおけるデータ表現を分析する方法が数多く提案されてきました。本研究では、脳における音や画像などの表現を調べるために用いられてきた方法が、人工ニューラルネットワークにも応用できるという可能性を示しました。

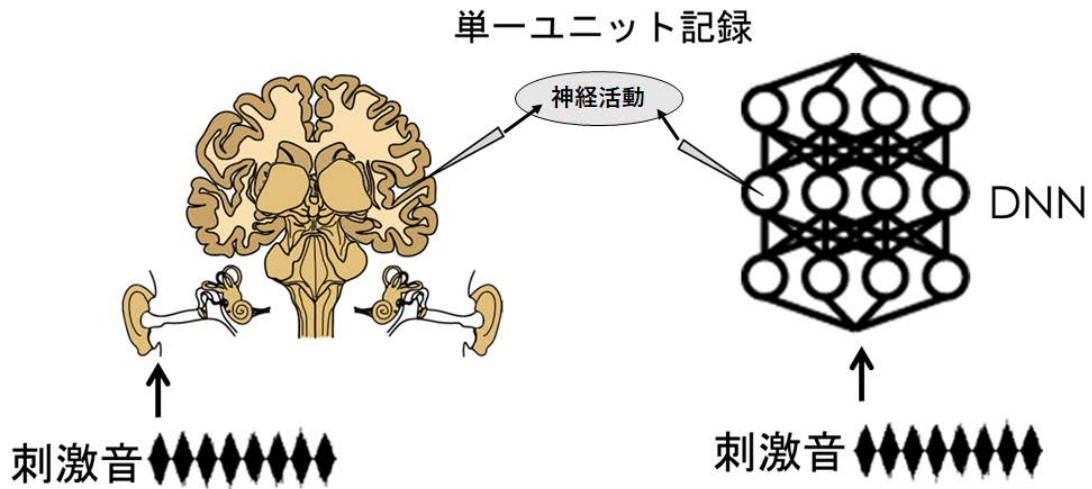


図 1: 本研究のアプローチ(模式図)。神経生理学実験では、脳内の個別神経細胞の活動を記録する(左)。本研究では聴覚系のモデルとして深層ニューラルネットワークを用いた(右)。自然な音認識課題に訓練させたニューラルネットについて、入力音に対する個別のユニット(素子)の反応を記録した。

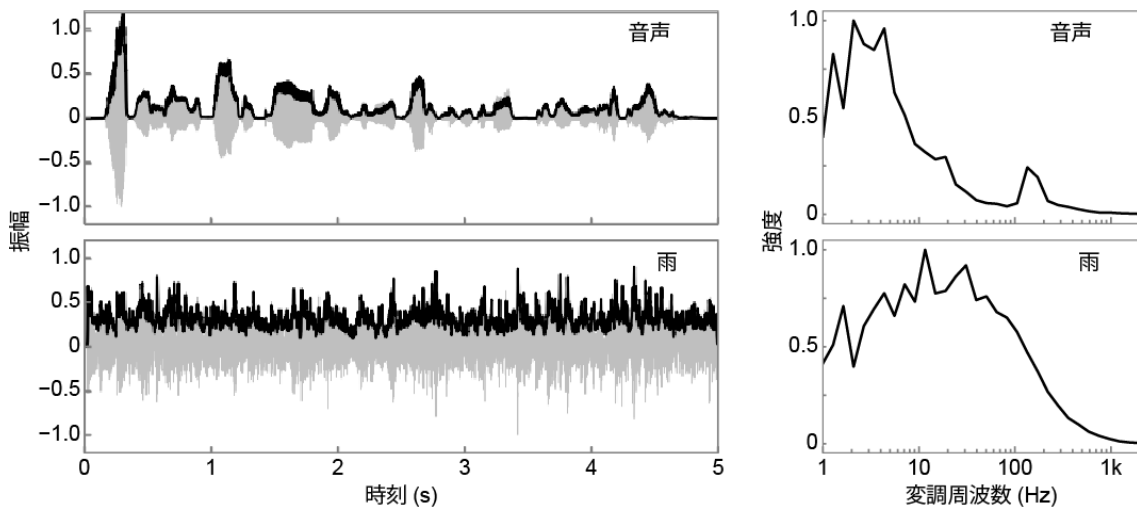


図 2: 音声と雨の音波形の例(左)とその変調スペクトル(右)。左図では音波形の瞬時振幅値(画面の分解能の限界のため、詳細な波形は見えない)を灰色で、その振幅値の包絡線を黒色で表す。振幅包絡線の変動を、振幅変調という。変調スペクトルは、振幅変調に含まれる周波数成分を表す。

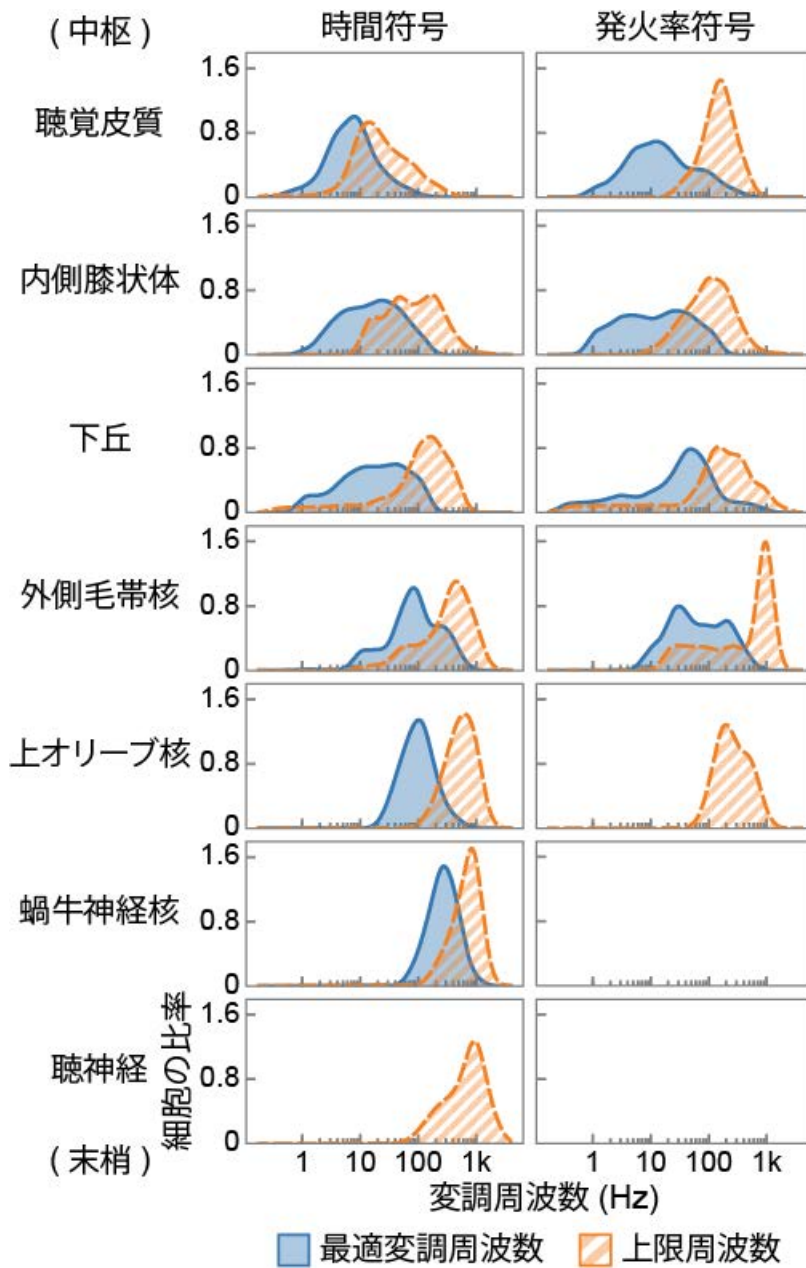


図 3: 聴覚神経系における神経細胞の最適変調周波数と上限周波数の分布。過去の神経生理学の研究のメタ分析により可視化した。時間符号では、最適変調周波数と上限周波数が、末梢から中枢にかけて徐々に低下する。発火率符号(平均発火率により変調周波数を表現)を示す細胞は、上オリーブ核や外側毛帯核より中枢の脳部位で発現する。

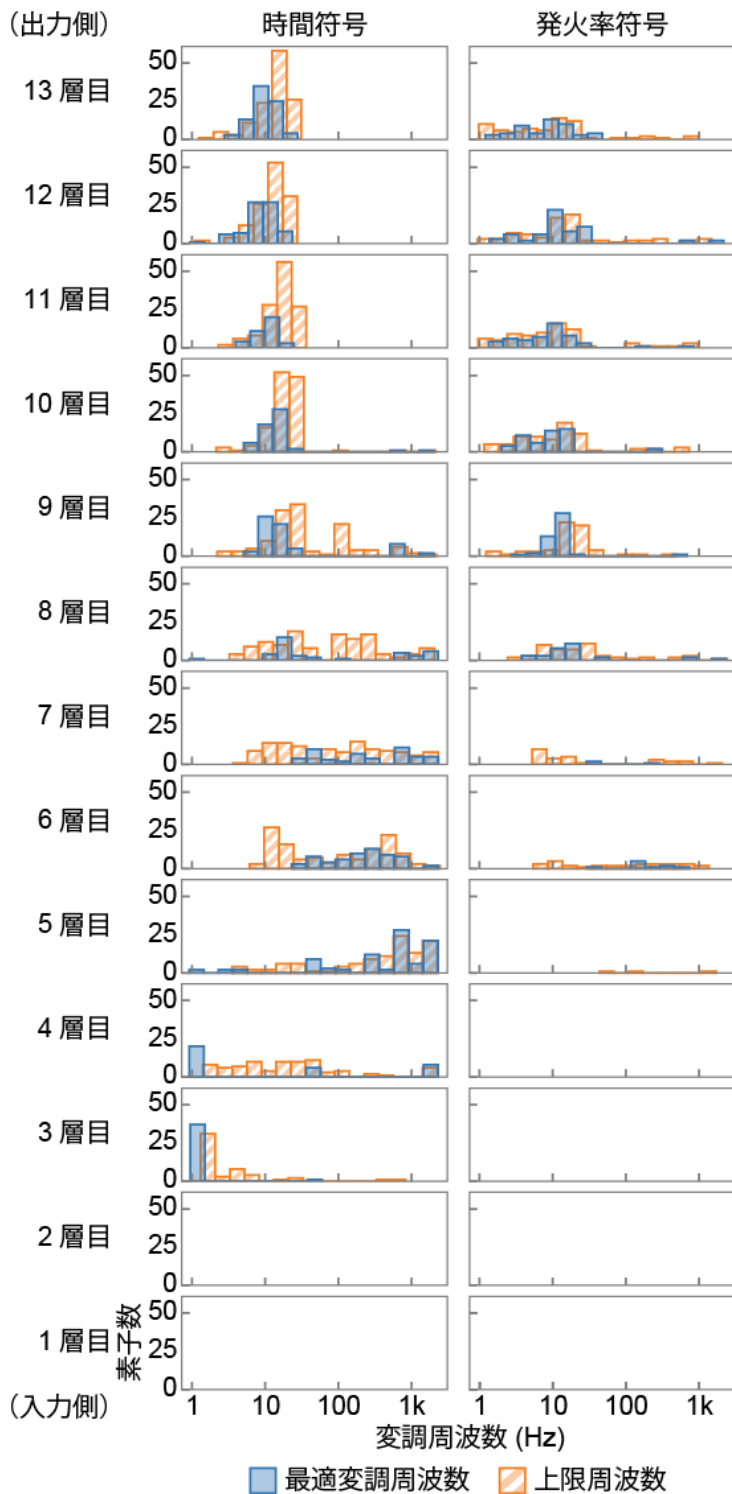


図 4: DNN における最適変調周波数と上限周波数の分布。時間符号では、中間から出力側の層にかけて、最適変調周波数と上限周波数が徐々に低下する。発火率符号については、中間の層より出力側の層において、平均活動強度により変調周波数が表現される。これらの傾向は、動物の聴覚神経系に見られる傾向(図2)と類似する。

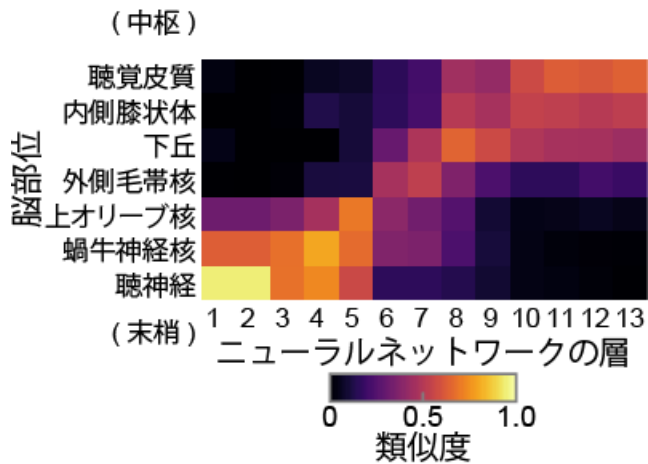


図 5: 聴覚系の各脳部位と DNN の層の間の類似度。ニューラルネットワークの入力側の層は末梢の脳部位と、中間の層は中間の脳部位と、出力側の層は中枢の脳部位と、それぞれ類似していた。

【用語解説】

※1 深層ニューラルネットワーク (DNN)

機械学習のモデルの一種である人工ニューラルネットワークの一つ。複雑な分類課題を高精度で行う有力技術としてもよく知られています。多数の素子からなる多数の層が縦列した構造により、データを処理します。ある層のある素子はその一段下の層の素子から入力を受け、単純な処理を経た後にその出力が次の層の多数の素子に伝達されます。このような構造は、元々は脳の神経細胞の結合様式に着想を得た技術です。その構造と機能が聴覚神経系と類似しているため、聴覚神経系の計算機シミュレーションにも適しています。本研究では、1つの素子を1つの神経細胞とみなして分析しました。

※2 振幅変調

音波形の振幅の緩やかな変化(振幅包絡)。変調の速さである変調周波数ごとの強度分布(=変調スペクトル)で表されることが多いです。振幅変調は音の認識に重要な音の特徴で、例えば、ヒトは、音の微細構造が壊されても、振幅変調のパターンが保たれていれば、その音をある程度認識することができます。内耳で音の周波数分析が行われるのと同様に、聴覚系では変調波形について何らかの周波数分析が行われるというのが通説です。

※3 聴覚神経系

音を認識する機能を持つ神経回路。耳にある内耳により、空気振動としての音信号が神経活動に変換され、その信号が聴神経に伝達されます。そこから多数の脳部位による段階的な処理を通して、信号が大脳皮質にある聴覚皮質に伝達されます。聴覚皮質では、さらに複雑な処理が行われ、音が認識されると考えられています。耳に近い部位は末梢、大脳皮質に近い部位は

中枢と呼ばれます。各脳部位は多数の神経細胞から構成されています。本研究では、末梢から中枢までの脳部位を、次のように分けました：聴神経、蝸牛神経核、上オリーブ核、外側毛帯核、下丘、内側膝状体、聴覚皮質。

※3 研究体制について

本研究は NTT 単独による研究成果です(共同研究ではない)。

平成 27～31 年度文部科学省科学研究費補助金新学術領域研究 多元質感知 (JP15H05915、JP16H01665)の助成を受けました。

<本件に関する問い合わせ先>

日本電信電話株式会社
先端技術総合研究所 広報担当

TEL046-240-5157