

fMRI を用いたヒト脳聴覚野からの音階デコードシステムの最適化

Optimization of sound scale decoding system from auditory cortex of human brain with fMRI

知能機械システム工学コース

知能メカトロダイナミクス研究室 1215007 繁本 成美

1. 緒言

脳機能画像解析法の発達により、脳神経科学の研究は飛躍的進歩を遂げている。特に、fMRI (functional Magnetic Resonance Imaging) は、脳機能画像解析法の代表格であり、空間解像度 3mm^3 と脳活動を細かく非侵襲的に測定することが可能なことから、近年、脳情報デコーディング技術に用いられている。中でも、視覚野の分野においては、被験者が見た物体だけでなく、夢で見た脳内イメージもデコーディング可能なデコーダアルゴリズムが開発されている⁽¹⁾。しかし、fMRI 撮像音が非常に大きいため、聴覚野の脳活動メカニズムも脳情報デコーディング技術も研究が遅れている。先行研究⁽²⁾では、15 度差間隔 4 音 (B2 (125Hz), B4 (500Hz), B6 (2kHz), B8 (8kHz)) を聴覚刺激として呈示した際の聴覚野の脳活動において、高い音程では聴覚野後方、低い音程で聴覚野前方が比較的強く反応したことが目視で確認され、聴覚野における周波数局在性が認められた。しかしながら、周波数解像度が低いという課題がある。

本研究では、fMRI の脳活動情報をデコーダで解析し、被験者が聴いた音程を推定する音階デコードシステム (図 1) の開発を行う。音階デコードシステムにおいて、デコーダが重要な要素であり、デコーダのアルゴリズムがシステムの精度に大きな影響を与える。そこで、本研究の目的を音階デコードシステムの最適なデコーダアルゴリズムを検討することとし、第一段階として、信号やパターン認識に強い Deep Belief Networks と画像認識に特化した Convolution Neural Network のアルゴリズムによる推定実験の結果を比較する。推定する刺激内容を C7 (2097Hz)・C#7 (2217Hz) の 1 度差とし周波数解像度を向上させ、聴覚野における脳情報デコーディング技術を開発する。

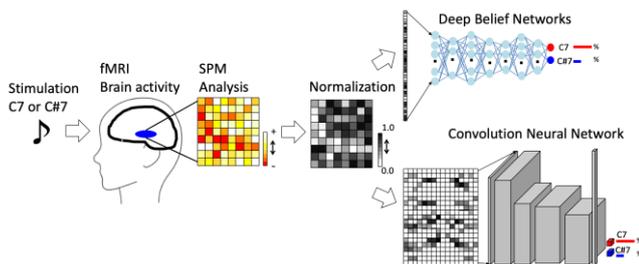


Fig. 1 Sound scale decode system flow

2. 脳機能画像撮像手法・解析手法

本研究では、SIEMENS MAGNETOM Prisma3T⁽³⁾ を使用し、脳活動を撮像した脳機能画像と脳の詳細な構造を撮像した構造画像を撮像する。脳機能画像と構造画像のデータは DICOM 形式であるため、NIFTI-1 形式に変換した後、画像解析に SPM12⁽⁴⁾ (Statistical Parametric Mapping) を用いて前処理および個人解析を行う。前処理では、撮像中の頭部の動きの補正 (Realignment)、脳スライス間の時間ズレの補正 (Slice Timing Correction)、構造画像への重ね合わせ (Coregistration)、

脳画像の空間的標準化 (Normalization)、空間的平滑化 (Smoothing) の処理を行っている。個人解析では、Session 内の同刺激の scan をランダムに複数抽出し onset としてコントラストを作成し統計解析を行う。本研究では、聴覚野における音階デコードシステムの開発を行うため、1 次聴覚野のみの脳活動を出力させるために関心領域 (Region of Interest: ROI) を側頭葉の上側頭回と横側頭回に位置する 1 次聴覚野 (Brodmann map⁽⁵⁾ 41&42) に設定する。

3. 深層学習手法

3.1 Deep Belief Networks⁽⁶⁾

本研究で使用する深層学習の 1 つは DBN (Deep Belief Networks) である。DBN は深層学習の手法の一種であり、教師なし学習である制約付きボルツマンマシン (Restricted Boltzmann Machine, 以下 RBM) と、教師あり学習であるロジスティック回帰 (Logistics Regression) から構成される (図 2)。DBN では多層に積み重ねられた RBM の各層で出力された特徴を次の層の RBM の入力とし、特徴抽出を行う。DBN の最後の層では、ロジスティック回帰を用い教師あり学習を行い効果的な学習を実現している。本研究で用いる DBN のハイパーパラメータは、各 RBM 層の細胞数である。

DBN を含むニューラルネットワークでは、学習を行う中で過学習を抑制するために、学習誤差に任意の閾値を設定し、各出力細胞の教師値に出力値が近似しないようにする。学習誤差 ε の算出は、式(1)により行う。

$$\varepsilon = \frac{1}{N} \sum_{k=1}^N \sum_{n=1}^{n_0} |T_n^k - O_n^k| \quad (1)$$

ここで、 N は総学習データ数、 k は RBM の各層での学習回数、 n_0 は出力細胞数、 T_n^k は n 番目の出力細胞の教師値、 O_n^k は n 番目の出力細胞の出力値である。図 2 に示す k は RBM の学習回数、 m はロジスティック回帰の学習回数である。

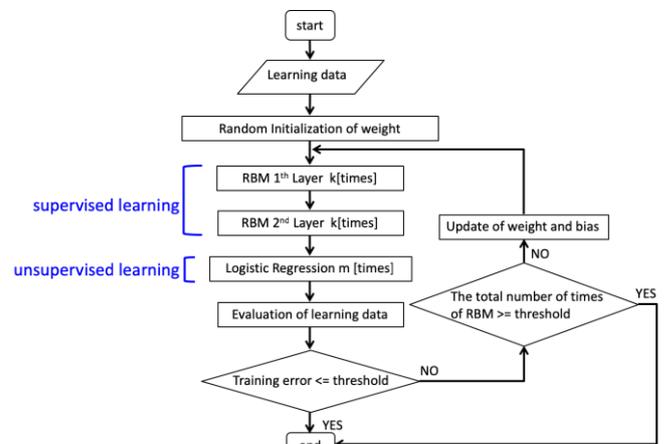


Fig.. 2 DBN flow

3.2 Convolution Neural Network (7)

本研究で提案する音階デコードシステムで最適なデコーダ検討で用いるもう 1 つの深層学習の手法は、CNN (Convolution Neural Network) である。全結合のニューラルネットワーク (NN) ではデータの形状情報がなく、NN が独自で個々のノードから特徴を抽出するのに対し、CNN は縦・横・チャンネル方向 (カラスケール) の 3 次元の画像から特徴を抽出し、学習・評価を行うため、画像認識に特化した深層学習手法である。

図 3 に本提案システムで使用する CNN の構造図を示す。CNN は、畳み込み層とプーリング層と呼ばれる 2 種類の層を交互に積み重ねた構造を持つフィードフォワード型の NN である。CNN の出力層直前には、隣接層間のノードすべてを結合した全結合層を 1 層配置し、出力層との結合は通常の NN と同様に設計する。本提案システムでは、クラス分類を目的とするため、目的のクラス数と同数のノードを配置し、活性化関数にソフトマックス関数 (softmax function) を用いる。

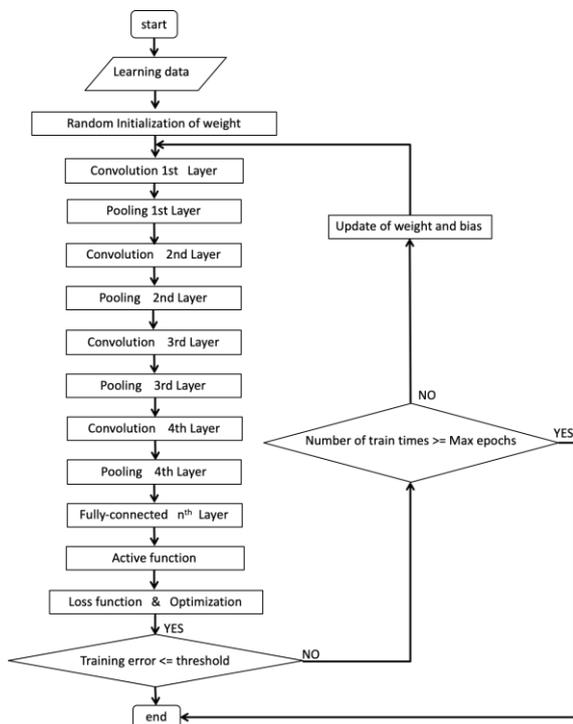


Fig. 3 CNN flow

3.3 評価方法 (8)

クラス分類問題では、各クラスの識別率(再現率)や全体の識別率(正確度)の他に、適合率と再現率の調和平均をとった F-値も利用されている。表 1 において、適合率とは、Positive と判定されたものの中に正解が Positive である割合を示し、再現率とは、正解が Positive であるものを Positive と判別できた割合を示し、適合率と再現率の間にはトレードオフの関係がある。両者の調和平均を F-値と呼び、F-値が高くなる場合が適合率・再現率のバランスが良い結果となる。表 1 における、再現率、適合率、F-値は、以下の式(2)-(4)で求められる。本研究では、識別実験の性能評価として、全体の識別率と各クラスの F-値の平均(全体 F-値)の値を用いる。

Table 1 Confusion matrix

		Predicted Class	
		Positive (P)	Negative (N)
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Recall = \frac{TP}{P} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

4. 実験

4.1 実験方法

本実験では、fMRI 撮像音の影響が少なく、可聴領域かつ聴力検査の範囲内である C7 と C#7 の 2 音を聴覚刺激として呈示し、そのときの脳活動データを取得する。聴覚刺激の呈示は、fMRI 撮像音の影響を低減する Active Noise Control 方式薄型ヘッドフォン OptoACTIVE⁽⁸⁾で行い、聴覚刺激呈示時間 9 秒、安静時間 15 秒を繰り返す図 4 のブロックデザインで行う。聴覚刺激は、C7・C#7 それぞれ 1 session につき 20 回ずつ呈示し、それを 3 session 行い、解析に用いるデータは、聴覚刺激呈示終了後の脳活動データとする。1 session ごとに C7・C#7 それぞれ刺激 4 回分の脳活動データをテストデータに、残りの 16 回分を学習データとする。

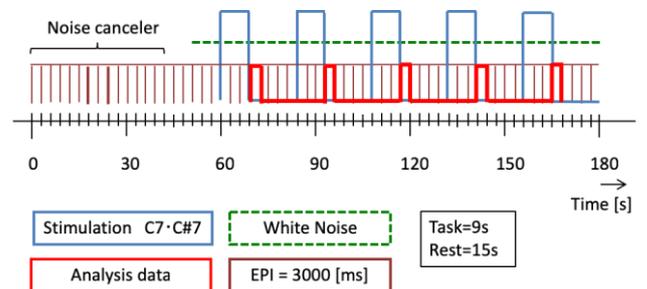


Fig. 4 fMRI experiment design

fMRI で取得した脳機能画像と構造画像に対して SPM12 を用いて 2 章で述べた前処理および個人解析を行う。脳の反応は一定ではないため、1 音につき、学習データ用脳活動データ (48 データ) から 4 scan onset の組み合わせを 4 種類用意し、48 コントラスト作成する。音階デコードシステムを、不安定な脳活動にも対応できるようにするため、テストデータ用脳活動データ (12 データ) から、1 scan onset で 12 コントラスト作成する。学習データ、テストデータ共に、1 次聴覚野すべてのボクセルから t 統計量を取得し、1 コントラスト毎に t 統計量を 0.0~1.0 の連続値に正規化を行う。DBN への入力データは、取得した t 統計量を 1 列に並べた txt 形式データとする。CNN への入力データは、ボクセルの位置情報から 100x48 行列に変換する。このとき、値を持たない位置の要素は 0.0 で埋め、csv ファイル形式とする。本実験には、簡易聴力検査において異常がみられない 20 代日本人健常男性 5 名 (絶対音感なし) の被験者が参加する。被験者には、高知工科大学倫理委員会承認に基づく fMRI 装置についての説明、および fMRI 実験の内容や安全性、個人情報保護についての説明をしたのち、書面による同意を得た上で実験を実施した。

本実験で識別する対象が C7・C#7 の 2 クラスであるため、出力クラスは 2 クラスとする。DBN は 3.1 節で述べたハイパーパラメータである細胞数を 100~800 の範囲内で 1st Layer \geq 2nd Layer となる組み合わせで実行する。CNN は、3.2 節で述べたハイパーパラメータである畳み込み層のチャンネル数を 16~64 の範囲内で 1st Layer \geq 2nd Layer \geq 3rd Layer \geq 4th Layer となる組み合わせで実行する。DBN, CNN ともに学習誤差が 0.1 以下になった場合に学習を終了し、評価へ移行する。本実験では、被験者毎に学習データとテストデータを作成し、DBN 識別実験と CNN 識別実験を行い、最もよい全体 F-値が得られるハイパーパラメータの組み合わせを探す。

4.2 実験結果

被験者毎にもっとも高い全体 F-値を得た DBN での識別実験結果を表 2 に、CNN での識別実験結果を表 3 に示す。DBN 識別実験の被験者 D は、最も良い全体 F-値が得られたハイパーパラメータ条件が 2 組存在し、表 2 に両方示している。

Table 2 DBN result

Subject ID	Layer1	Layer2	C7/(12)	C#7/(12)	Total/(24)	Total(%)	F-measuer
Subject A	700	500	9	7	16	66.67%	0.6643
Subject B	300	200	7	8	15	62.50%	0.6243
Subject C	300	100	6	9	15	62.50%	0.6191
Subject D	200	100	7	8	15	62.50%	0.6243
	400	200	8	7			
Subject E	300	100	4	10	14	58.33%	0.5556

Table 3 CNN result

Subject ID	Filter size	Number of channel				C7/(12)	C#7/(12)	Total/(24)	Total(%)	F-measuer
		1st Layer	2nd Layer	3rd Layer	4th Layer					
Subject A	2	32	64	64	64	9	9	18	75.00%	0.7500
Subject B	3	8	8	8	16	6	9	15	62.50%	0.6190
Subject C	5	8	16	16	16	8	9	17	70.83%	0.7078
Subject D	2	32	32	32	32	7	8	15	62.50%	0.6243
Subject E	2	8	16	16	16	3	9	12	50.00%	0.4667

表 2, 3 より、被験者 E 以外の 4 名において、DBN より CNN の方が識別率・全体 F-値が高い結果が得られており、DBN で識別率・全体 F-値が高い被験者は、CNN でも高い結果が得られていることがわかる。したがって、入力データにおいて、深層学習で学習がしやすいデータであったことに加え、ハイパーパラメータが適切な値に近い値で設定できていると考えられる。

表 2 より、DBN では、ハイパーパラメータである RBM2 層の細胞数が被験者 5 名中 4 名が 300 以下で 60%前後と同程度の結果が得られている。被験者 A でも RBM2 層の細胞数が [300, 200] のとき、識別率 62.50%、全体 F-値 0.6243 と 4 名の被験者と同程度の結果であったことより、DBN では、異なる被験者でも RBM2 層の細胞数が 300 以下程度で約 60%の識別率を得ることができると推測できる。

CNN のハイパーパラメータは、畳み込み層のフィルターサイズとチャンネル数である。表 3 より、被験者ごとに高い識別率・全体 F-値が得られるハイパーパラメータは異なり、関係性がみられなかった。被験者 E では、どのハイパーパラメータの組み合わせにおいても識別率が 50%前後であり、今回のハイパーパラメータ条件における識別率への影響は小さいことが推測できる。したがって、全被験者で識別可能な CNN システムにするためには、CNN の構造や使用する活性化関数をより適切に設定することが必要である。

5. 結言

本研究では、fMRI で測定した 1 度差の聴覚刺激を受けたヒト脳の聴覚野の脳活動から、被験者がどのような純音を聴いているか DBN, CNN を用いた提案手法で識別を行った。結果、DBN より CNN の方が、提案する音階デコードシステムに適しており、最大識別率は 75.00%、全体 F-値は 0.750 と既報⁽⁶⁾での最大識別率 72.5% より高い結果を得ることができた。要因の一つとして、刺激に対して各聴覚野内の位置 (Voxel) において脳活動の微少な変化が影響していると考えられる。DBN は個々の Voxel における脳活動のみから学習するのに対し、CNN は注目する Voxel の前後左右に存在する Voxel の脳活動も含めて学習するため、DBN より脳活動の変動に強いアルゴリズムであると言え、DBN より CNN の方が識別率・全体 F-値が高くなったと考えられる。

しかしながら、識別率が低い被験者も存在し、CNN は、構造や活性化関数の見直し、DBN は、より適切なハイパーパラメータ設定が求められる。

文献

- (1) Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y, "Neural decoding of visual imagery during sleep", Science, 340(6132), 639-642. (2013)
- (2) Dave R.M.Langer, Walter H.Backes, and Pim van Dijk: "Representation of lateralization and tonotopy in primary versus secondary human auditory cortex", Neuro Image34, pp264-273, (2007)
- (3) fMRI <https://www.healthcare.siemens.co.jp/magnetic-resonance-imaging/research-systems/magnetom-prisma> (2019 年)
- (4) SPM <http://www.fil.ion.ucl.ac.uk/spm/> (2019 年)
- (5) 菊池吉晃, 妹尾淳史, 安保雅博, 渡邊修, 米本恭三, (2012), "SPM8 脳画像解析マニュアル fMRI, 拡散テンソルへの応用"
- (6) 繁本成美, 佐藤公信, 芝田京子, 井上喜雄, 園部元康: "深層学習を用いた聴覚野からの音階デコードシステムにおける周波数解像度の向上", 日本機械学会 中四国支部第 55 期卒業論文発表会, 講演番号 914, (2017)
- (7) 繁本成美, 佐藤公信, 芝田京子, 井上喜雄: "CNN を用いたヒト脳聴覚野における音階デコードシステムの実装", 計測自動制御学会四国支部学術講演会 2018, 講演番号 PS2-15, (2018)
- (8) OptoACTIVE <http://www.optoacoustics.com/medical/optoactive-ii> (2019 年)
- (9) 平井有三, (2012), "はじめてのパターン認識"