

fMRI を用いたヒト脳聴覚野からの和音における特定音検出

システム工学群

ヒューマンメカトロダイナミクス研究室 1200071 新家 潤

1. 緒言

近年、fMRI (functional Magnetic Resonance Imaging) の発達により脳情報デコーディング技術の研究が進められている。fMRI は脳機能画像解析法の代表格であり、脳活動を定量的に測定することが可能である。脳情報デコーディング技術のなかでも、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、脳科学における聴覚野の研究分野では、MRI 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。

そこで、本研究グループでは、fMRI の脳活動情報を解析し、被験者が聴いた音を推定する音階デコードシステムの開発を目的としている。Langer らの先行研究⁽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、高い音階では聴覚野後方、低い音階で聴覚野前方が強く反応したことが確認され、聴覚野における周波数局在性が認められた。既報⁽³⁾では、周波数解像度を上げることを目的に、音階デコードシステムのデコーダアルゴリズムとして CNN (Convolution Neural Network) と DBN (Deep Belief Network) を用いて、1 音階差 (C7 (2097Hz), C#7 (2217Hz)) の音の推定実験を行った。20 代男性 3 名を対象に推定実験を行ったところ、CNN のアルゴリズムを用いることで最大識別率 75.00%、平均 69.44% が得られ、単音であれば検出可能である見通しが得られた。この実験結果を受けて、本研究では既報⁽³⁾の音階識別システム SSIS (Sound Scale Identification System) を応用することで日常生活にある様々な音階の混ざった複雑な音を CNN を用いて脳画像から推定できると考え、和音を対象とし和音の中に含まれる音の検出を目的とする (図 1)。本報では第一段階として、三和音からの特定の 1 音 C7 の検出を行う。

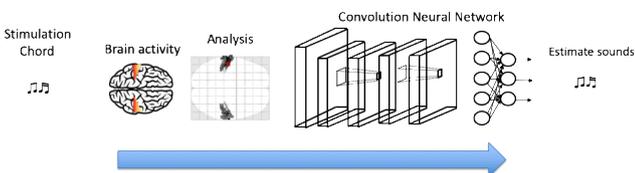


Fig. 1 Sound scale decode system flow

2. 脳機能画像撮像手法・解析手法

本研究では、既報⁽³⁾の脳機能画像撮像手法と解析手法を用いる。fMRI (SIEMENS MAGNETOM Prisma3T) を使用し、聴覚刺激に対する脳機能画像と脳の構造画像を撮像し、DICOM 形式で保存する。脳機能画像と構造画像のデータを 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を補正し (Realignment)、脳スライス間の時間ズレを補正 (Slice Timing Correction) し

たあと、構造画像への重ね合わせ (Coregistration) をして、脳画像の空間的標準化 (Normalization)、および、空間的平滑化 (Smoothing) の処理を行う。このデータを基に個人解析では、同刺激のデータをランダムに複数選択し、統計解析を行う。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 (Broadman map 41&42) に設定し 1 次聴覚野のみの t 統計量を取得する。t 統計量は 1 コントラスト毎に取得し、各コントラスト内で 0.0~1.0 の連続値に正規化した数値を CSV 形式で出力し、位置情報をもとに深層学習で用いる H100xW48 の入力データに変換する。

3. 特定音検出手法

本研究では、前章で作成したデータから特定音を検出する技法として既報⁽³⁾の SSIS を用いる。SSIS は深層学習の手法として、画像認識に特化した学習手法である CNN を用いて、学習と評価を行う。CNN の構造は、畳み込み層とプーリング層を交互に配置し、畳み込みとプーリングを行う。畳み込みとプーリングによって得られた特徴から出力を得るために、CNN の出力層直前に全結合層 1 層を配置している。出力層との結合は NN (Neural Network) で設計、クラス分類を目的とし、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) を用いて出力を確率値に変換し分類する。

4. 実験

4.1 実験方法

本実験は、既報⁽³⁾の実験モデルを参考にした。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⁽⁵⁾で聴覚刺激の呈示を行い、fMRI の動作音の影響を抑えた。OptoACTIVE のノイズを低減する機能 ANC (Active Noise Control) を有効に使うには 40 秒程度の時間がかかるため、実験開始 1 分間を ANC の較正時間とし、実験開始 1 分後の刺激から脳機能画像を取得する。fMRI の機能画像撮像パラメータを表 1 に示す。刺激音は 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音 C7 を基準に 1 オクターブ選択した。聴覚刺激に用いる刺激音は C7, D7 (2349Hz), E7 (2637Hz), F7 (2793Hz), G7 (3135Hz), A7 (3520Hz), B7 (3951Hz) の 7 音で、7 音の中からランダムに 3 音選択した和音を聴覚刺激として呈示し、そのときの脳機能画像を取得する。

Table 1 Imaging parameters of structural images

Echo time (TE)	48 [ms]
Repetition time (TR)	3000 [ms]
Field of view (FOV)	192 × 192 [mm]
Filip angle	90 [°]
Matrix size	2.0 × 2.0 × 3.0 [mm]
Slice thickness	3.0 [mm]
Slice gap	0.75 [mm]
Slice	36 [slices]
Slice acquisition order	Ascending

聴覚刺激の呈示は、聴覚刺激と安静を繰り返すブロックデザインで行う (図 2)。本研究で用いる聴覚刺激は、Chord 1 (C7を含む和音) と Chord 2 (C7を含まない和音) の 2 条件で行う。刺激は 9 秒間呈示し、刺激と刺激の間には 9 秒間の安静時間を設ける。被験者の実験時間増加の負荷を考慮し各条件 60 回ずつの刺激呈示を 1Session あたり 20 回として 3Session 行う。それぞれ刺激 60 回分の脳活動データをランダムに評価データ (12 データ)、学習データ (48 データ) に分ける。

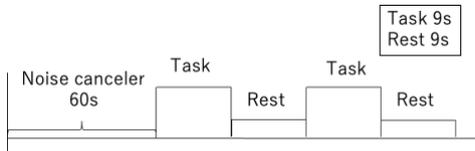


Fig. 2 fMRI experiment design

fMRI で取得した脳機能画像を、2 章で述べたように SPM12 を用いて前処理および個人解析を行う。学習用のデータ (48 データ) から session 毎に 4scan の組み合わせを 4 セット用意し、4scan×4set×3session=48 コントラスト作成する。評価用のデータ (12 データ) は、1 scan × 1set × 3session=12 コントラスト作成する。このデータを用いて 3 章で述べたように CNN を用いて学習および評価を行う。CNN の学習条件を表 2 に示す。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。このデータをもとに C7 が含まれた和音か否かを識別し、C7 を聴いたときの脳活動を検出できたかどうか判断する。

Table 2 CNN learning conditions

			Set value
Condition	Number of CNN layers		4
	Convolution	stride	1
		filter size	2
	Pooling	stride	2
		filter size	2
	Learning rate		0.0001
	Drop out		0.7
Error rate		0.1	
Termination condition			erro rate
Hyper parameter	Convolution	Filter size	2~64(exponent of 1)
		Channels	2~64(exponent of 2)

あらかじめ本実験で用いる 7 音を聞かせて異常がみられなかった 20 代日本人健康男性 6 名の被験者が実験に参加した。実験は、本学倫理委員会の承認を得た上で、被験者に内容を説明し同意を得て実施した。

4.2 実験結果と考察

被験者毎にハイパーパラメータ値を試行錯誤で設定し複数回試行した。被験者毎に最も高い識別率を得たときのハイパーパラメータの設定値と識別率を表 3 に、最も低い識別率を得たときのハイパーパラメータの設定値と識別率を表 4 に示す。表 3 より、最高識別率を得たときのハイパーパラメータ

Table 3 Maximum identification result

Subject ID	Hyper parameter					Results			
	Filter size	Number of channel				C7 (/12)	C#7 (/12)	Total (/24)	Total
		1st Layer	2nd Layer	3rd Layer	4th Layer				
A	16	32	32	32	32	7	10	17	70.83%
B	8	4	4	4	4	6	7	13	54.17%
C	8	4	4	4	4	8	6	14	58.33%
D	2	4	4	4	4	5	10	15	62.50%
E	8	2	2	2	2	5	10	15	62.50%
F	8	8	8	8	8	7	7	14	58.83%

タの設定値は、被験者によって異なった。また、同じハイパーパラメータの設定値であっても、表 3 より被験者 A では最高識別率、表 4 より被験者 F では最小識別率であったように、ハイパーパラメータの適値は被験者によって異なった。加えて、表 3 と表 4 より、どの被験者においても、識別率の最高値と最低値の差が 20%前後あり、ハイパーパラメータの設定値によって識別率が大きく変動した。

また、表 3 より、被験者 A では最大の 70.83%の識別率が得られた。本報で参考とした SSIS の単音における識別結果の最大 75.00%、平均 69.44%の識別率と比較すると、被験者 A では平均識別率を上回っており、C7 の特徴を検出できているといえる。よって、fMRI の撮像音の影響の少ない音であれば、3 音の中から C7 を特定音として検出できる見通しが得られた。しかしながら、被験者 A 以外の識別率は既報³⁾の平均識別率を大きく下回っている。

識別率が低い原因として、学習データ数が少ないことがあげられ、被験者に負担の少ない実験データ取得の方法を今後新たに提案したい。また、前述したように、ハイパーパラメータの設定値が大きく識別率に影響を与えたことが考えられる。今回、どの被験者の識別実験においても複数回試行錯誤でハイパーパラメータの適値を探索したが、試行した範囲外に最適値が存在する可能性は否めない。したがって、識別率の向上には、最適なハイパーパラメータの設定と学習データの数を増やすことが考えられる。

5. 結言

本研究では、fMRI を用いて三和音を聴覚刺激として呈示した際の脳活動から特定音 C7 の検出を行った結果、全体としては識別率が低かったものの被験者 A においては最大識別率 70.83%という結果が得られた。被験者 A も含めさらなる識別率向上が必要であるが、実験結果より、本提案法によって C7 以外の音であっても識別可能である見通しが得られた。今後は、実験データ取得方法の模索とハイパーパラメータの最適化によりさらなる識別率向上を考えたい。

文献

- (1) Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y, "Neural decoding of visual imagery during sleep", Science, 340(6132), (2013), pp.639-642.
- (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, (2007), pp264-273.
- (3) Narumi Shigemoto, Hironobu Stoh, Kyoko Shibata, and Yoshio Inoue, "Study of Deep Learning for Sound Scale Decoding Technology from Human Brain Auditory Cortex", 2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech), (2019), pp212-213.
- (4) SPM <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/> (2019/01/17).
- (5) OptoACTIVE <http://www.optoacoustics.com/medical/optoactive/features> (2019/01/17).

Table 4 Minimum identification result

Subject ID	Hyper parameter					Results			
	Filter size	Number of channel				C7 (/12)	C#7 (/12)	Total (/24)	Total
		1st Layer	2nd Layer	3rd Layer	4th Layer				
A	32	32	32	32	32	2	11	13	54.17%
B	2	2	2	2	2	5	4	9	37.50%
C	2	2	2	2	2	5	4	9	37.50%
D	4	4	4	4	4	4	5	9	37.50%
E	16	16	16	16	16	6	2	8	33.33%
F	16	32	32	32	32	3	5	8	33.33%