

2021 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

生体情報と機械学習による楽曲聴取者の感情推定

Estimating Music Listener's Emotions
by Using Bio-Signals and Machine Learning

谷澤 七海
(5120F049-7)

提出日：2022.1.24

指導教員：亀山渉教授

研究指導名：マルチメディア情報流通システム研究

目次

第 1 章 序論.....	1
1.1 研究背景.....	1
1.2 研究目的.....	1
1.3 本論文の構成.....	2
第 2 章 関連研究.....	3
2.1 基礎律動.....	3
2.2 心拍間隔(RRI).....	3
2.3 瞳孔径.....	4
第 3 章 実験.....	5
3.1 実験概要.....	5
3.1.1 被験者.....	5
3.1.2 実験装置.....	5
3.1.3 使用した楽曲.....	5
3.1.4 アンケート.....	10
3.1.4.1 聴取後の評価.....	10
3.1.4.2 聴取中の評価.....	10
3.2 実験手順.....	11
第 4 章 解析データと解析方法.....	13
4.1 分類による解析.....	13
4.1.1 聴取後と聴取中の評価を使用した場合の比較.....	13
4.1.1.1 データ整形.....	13
4.1.1.2 正解ラベルの作成.....	15
4.1.1.3 解析手法.....	15
4.1.2 楽曲情報を追加した解析.....	17
4.1.2.1 データ整形.....	18

4.1.2.2	正解ラベルの作成.....	18
4.1.2.3	解析手法	18
4.1.3	AutoEncoder によるデータ拡張を使用した解析	18
4.1.3.1	データ整形.....	18
4.1.3.2	正解ラベルの作成.....	18
4.1.3.3	解析手法	19
4.2	回帰による解析.....	21
4.2.1	データ整形.....	21
4.2.2	正解ラベルの作成	21
4.2.3	解析手法.....	21
第 5 章	結果と考察	23
5.1	分類による解析.....	23
5.1.1	聴取後と聴取中の評価を使用した場合の比較.....	23
5.1.2	楽曲情報を追加した解析	25
5.1.3	AutoEncoder によるデータ拡張を使用した解析	25
5.2	回帰による解析.....	26
第 6 章	結論	29
6.1	まとめ	29
6.2	今後の課題	30
図一覧	31
表一覧	32
付録	33
参考文献	41
謝辞	44
研究業績	45

第1章 序論

1.1 研究背景

近年、ストリーミングサービスの普及により様々な音楽コンテンツが流通するようになり、その量は膨大なものとなっている。このことにより、ユーザにとって求めている音楽コンテンツを探すことが難しくなっていると考えられる。これに対する解決策として、推薦システムがあり、現在使われている推薦システムでは再生履歴等をもとに推薦を行なっている。しかし、この現在の推薦システムでは同じ音楽コンテンツであってもユーザによって抱く印象や感情は異なるためユーザの嗜好を必ずしも反映できているとは限らない。そこで、音楽聴取時において生体情報を計測し、その生体情報をもとにしたコンテンツ推薦システムが検討されている。生体情報から推定される聴取楽曲に対する感情や印象をメタデータとして付与することで、よりユーザの嗜好にあったコンテンツ推薦ができると考える。

1.2 研究目的

まず、これまでの研究では、音楽聴取時に得られた聴取者の生体情報を様々な方法で解析し、クラスタリングすることにより、聴取楽曲に対する好き嫌いを分類していた。先行研究[1]では、得られた生体情報をサポートベクターマシンにより解析することで、外挿による検証(学習と検証で異なる楽曲の生体情報を使用する検証方法)でチャンスレベルを超える精度が得られたとされている。また、先行研究[2]では、Variational AutoEncoder で解析し、特徴表現をクラスタリングした結果、特徴のあるクラスタが存在するとされている。これらのことから生体情報を用いることによって楽曲に対する「好き」「嫌い」は推定可能であると推測できる。次に、以上を踏まえて、楽曲に対して抱いた感情や印象で細かく判別するための検証として、文献[3]に示す実験を新たに行い、「楽しい」、「悲しい」、「イライラする」、「落ち着く」の4種類の感情を3種類の生体情報(基礎律動、心拍間隔、瞳孔径)から推定するモデルの作成を行った。この結果として、10人の被験者に対し、内挿(学習と検証で同じ楽曲の生体情報を使用する検証方法)による検証ではチャンスレベルを超える精度となり、生体情報から感情を推定できる可能性が示唆された。しかし、一方で、解析に使えるデータ量が少ないため外挿による検証を行っておらず、内挿による検証において推定精度が高くてもモデルに汎化性があるかがわからないという課題があった。さらに、この検証では、1曲ごとに感情ラベルが1つ割り当てられているが、楽曲内でも感情が変化する可能性があるため、この点を考慮できていないという課題もあった。そこで、これらの課題に対し、本論文ではまず、取得データの量を増やすために被験者を筆者自身とした新実験を行い、外挿による検証を行なった。また、文献[3]で取得していた楽曲聴取後の評価に加え、楽曲聴取中もジョイスティックを用いて評価を取得し、これを使用して楽曲内での感情の変化を考慮した検証も行なった。さらに、外挿による推定精度向上のために

Spotify[4]より取得した楽曲情報を利用した検証や **AutoEncoder** を用いたデータ拡張を追加した検証も行った。

1.3 本論文の構成

本論文は以下の 6 章で構成されている。

第 1 章では、本研究の背景、本研究の目的、及び本論文の構成を示す。第 2 章では、生体情報に関する先行研究と生体情報を用いた分類に関する先行研究を示す。第 3 章では、本研究にて行なった実験の概要について示す。第 4 章では、本研究にて行なった各検証にて使用したデータと解析方法について示す。第 5 章では、各検証における解析結果とそれに対する考察を示す。第 6 章では、本研究のまとめと今後の課題について述べる。

第2章 関連研究

2.1 基礎律動

脳波とは、時間経過により変化する脳の自発的電氣的活動を頭皮上の電極から記録したものである。脳波は波から構成されており、リズムを形成するが、このように一定の周波数の波が連続すると律動的に見える。このような、脳波の大部分を形成する特定の脳波活動を基礎律動という。脳波における周波数成分は α 波(8~13[Hz]), β 波(14~30[Hz]), θ 波(4~7[Hz]), δ 波(0.5~3[Hz]), γ 波(31[Hz]以上)に区分される[5][6]。特に、 α 波、 β 波に関して、これらに区分される周波数の脳波は人間のリラックス状態や精神活動状態により変動するとされており、 β 波と α 波の比率を評価することで脳の活性度や活動度を測ることができるとされている[6]。また、文献[7]に示す論文では、電極の対称性を利用し、各対称となっている電極で左右差、左右比を求め、これを解析に利用することで音楽を聴いた際の高精度な感情分類ができるとなっている。

2.2 心拍間隔(RRI)

心電図における基本波形は、図 2.1 に示すように、心房の興奮過程を示す P 波、心室の興奮過程を示す QRS 波、心室の興奮の消退過程を示す T 波により構成されている。R 波というのは心室の興奮過程を示す QRS 波のうち、上向きの波のことであり、この R 波の間隔を心拍間隔(RRI)と呼ぶ[8]。この心拍間隔のゆらぎは安静時に大きく、ストレス時には低減するとされている[9]。

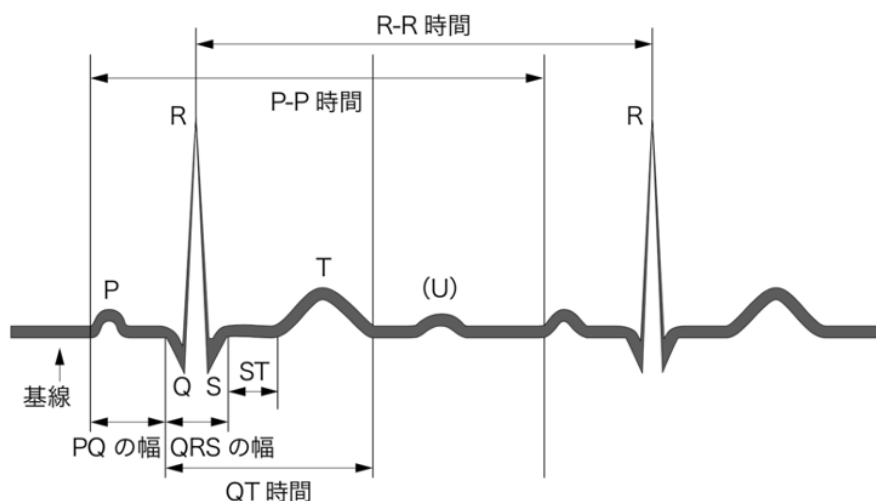


図 2.1 心電図における基本波形(文献[8]の図 2 より引用)

2.3 瞳孔径

瞳孔に関して、基本的には網膜に入射する光量を制御する役割があり、明所では瞳孔径が収縮し、暗所では瞳孔径が拡大する。しかし、このほかにも知覚や記憶といった認知プロセスとも瞳孔は関係があることが示唆されている[10]。瞳孔は、自律神経に支配されており、驚いたときや興奮したときなど交感神経が優位であるときは瞳孔径が拡大し、眠いときや疲れているときなど副交感神経が優位であるときは瞳孔径が縮小すると言われている[11]。また、瞳孔径は音楽の嗜好の推定に関する研究でも用いられている。文献[12]に示す論文では、瞳孔径を利用し、音楽の好みと馴染み度合いを推定しており、被験者による評定と推定した結果が80%以上の試行で近い値となるという結果が得られている。

第3章 実験

3.1 実験概要

3.1.1 被験者

被験者は23歳女性1人であり、筆者自身である。

3.1.2 実験装置

本研究では、基礎律動、心拍間隔(RRI)、瞳孔径を測定した。基礎律動の計測にはEMOTIV社製脳波計EPOCX(14チャンネル, 128[Hz]計測)を使用した。脳波計のチャンネルの位置は図3.1に示す通りである。図3.1において、青で示すものが計測する電極の位置となっており、赤で示すものが基準となる電極の位置となっている。RRIの測定にはPolar社製H10心拍センサ及び同社製GPSスポーツウォッチV800を使用した。瞳孔径の測定にはTobii Technology社製Eye Tracker Tobii X60(60[Hz]計測)を使用した。

また、楽曲聴取にaudio-technica社製ノイズキャンセリングイヤホンATH-ANC23を、聴取中の評価の取得にPXN Electronic Technology社製フライトシミュレーションジョイスティックコントローラーPXN-2113-SE(20[Hz]計測)を使用した。

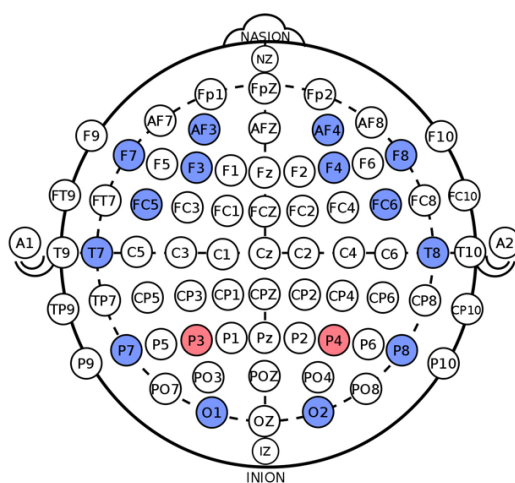


図 3.1 EPOCX の電極の位置(文献[13]の図より引用)

3.1.3 使用した楽曲

実験には被験者が選定した様々な印象の楽曲, 200曲(平均再生時間 249.12 秒, 標準偏差 45.77)を用いた。これら 200 曲の一覧を表 3.1~表 3.4 に示す。また、全ての楽曲について歌詞付きであり、歌詞は主に日本語で書かれたものとなっている。

表 3.1 実験で使用した楽曲の一覧(1~50 曲目)

音源番号	楽曲名	アーティスト名
1	Moondrop	aint
2	Stella	AllyourDays
3	Work	BBHF
4	さざんかの宿	大川栄策
5	Sonic Fireworks	avengers in sci-fi
6	真剣SUNSHINE	Hey! Say! JUMP
7	Party	AllyourDays
8	ハジメテノオト	malo
9	津軽海峡・冬景色	石川さゆり
10	Shout	androp
11	Anemia	a crowd of rebellion
12	初期衝動	BRAHMAN
13	DAYBREAK FRONTLINE	Orangestar
14	打上花火	DAOKO×米津玄師
15	HOOOWLL	Gero×ARAKI
16	デーモンダンストークョー	Eve
17	フライングゲット	AKB48
18	絶体絶命	Cō shu Nie
19	インドア系ならトラックメイカー	Yunomi&nicamoq
20	唯我独尊 ONLY ONE	CTS
21	助演男優賞	Creepy Nuts
22	キミの隣	halca
23	雨天決行	GOOD ON THE REEL
24	Easy Breezy	chelmico
25	Everyday、カチューシャ	AKB48
26	LION	坂口有望
27	Masquerade	SHE'S
28	Name of Love	cinema staff
29	neoromantic	江沼郁弥
30	Strangers	ARAKI
31	You Only Live Once	YURI!!! on ICE feat. w.hatano
32	スーパーヒーロー	Guiano
33	スパイス	Perfume
34	バナナフィッシュの浜辺と黒い虹(Yuuki Ozaki ver.)	Galileo Galilei
35	ベノム[MARETU Remix]	かいりきベア
36	ユメマカセ	SOIL&"PIMP"SESSIONS feat. Yojiro Noda
37	人生いろいろ	島倉千代子
38	誰でもないモンスター	夏と彗星
39	チューリングラブ feat. Sou	ナナヲアカリ
40	Happy… Good day!(feat. ド葛本社)	DJ Genki
41	VS	BLUE ENCOUNT
42	ハッピーシンセサイザ(feat. 巡音ルカ&GUMI)	EasyPop
43	パラダイムボックス	式島律(沢城千春)、佐竹笙悟(武内駿輔)
44	愛憎	CIVILIAN
45	怪獣の花唄	Vaundy
46	喰病しのアイデア	After the Rain
47	丸ノ内サディスティック	椎名林檎
48	軽蔑と始まり	DIRENGREY
49	突破口	SUPER BEAVER
50	籠の中のぼくらは	AIKI from bless4

表 3.2 実験で使用した楽曲の一覧(51~100 曲目)

音源番号	楽曲名	アーティスト名
51	オープニング「火事と喧嘩は江戸の華」 feat. KREVA, MIYAVI	石川さゆり
52	男気・夢・音頭	BOYS AND MEN
53	蜃気楼	10-FEET
54	A prisoner in the glasses	ぼくのりりっくのぼうよみ
55	ADORE	lynch.
56	Freesia	(K)NoW_NAME
57	Making it Drive	UVERworld
58	STARGAZER	ARAKI, nqrse
59	カゲロウデイズ	じん
60	シグナル	WANIMA
61	フィクサー	ぬゆり
62	リップサービス	The Floor
63	宇宙	CHRONICLE
64	仮面	タッキー&翼
65	群青サバイバル	小松未可子
66	蝶々結び	Aimer
67	北酒場	細川たかし
68	本能	椎名林檎
69	未来は誰かの手中	電波少女
70	恋するフォーチュンクッキー	AKB48
71	NON TiE-UP	BiSH
72	Tell Your World	livetune
73	カルベ・ディエム	The Charm Park
74	シグナル	ゲーム実況者わくわくバンド
75	シンデレラ・シンドローム	ヒステリックパニック
76	たのしいせいかつ	ぼくのりりっくのぼうよみ
77	よふかしのうた	Creepy Nuts
78	空洞空洞	amazarashi
79	川の流れるように	美空ひばり
80	BLUE feat. Kojikoji	LUCKY TAPES
81	Dance in the rain	Rude-a
82	Plyers	BRADIO
83	Lady	CTS
84	LOVE DOLL	DECO*27
85	Sunrise(re-build)	ぼくのりりっくのぼうよみ
86	オドルヨル	WANIMA
87	その向こうへ	10-FEET
88	ノンファンタジー	LIP×LIP
89	ハッピーエンド	back number
90	ミステリーヴァージン	山田涼介
91	ピルグリム	八十八ヶ所巡礼
92	雨の慕情	八代亜紀
93	四季折々に揺蕩いて	After the Rain
94	水泡	Ivy to Fraudulent Game
95	星の王子さま	ナユタン星人
96	天城越え	石川さゆり
97	曖昧サイボーグ(feat. うごくちゃん& JungMato)	Neko Hacker
98	MUSIC VIDEO	岡崎体育
99	水槽	中島愛
100	東京フラッシュ	Vaundy

表 3.3 実験で使用した楽曲の一覧(101~150 曲目)

音源番号	楽曲名	アーティスト名
101	7日間.	phatmans after school
102	Q	女王蜂
103	インドの牛乳屋さん	50TA
104	ヘビースモーク	にしな
105	遺失物取扱所	Suck a Stew Dry
106	夏を待っていました	amazarashi
107	救難戦争	HOWL BE QUIET
108	命に嫌われている	Sou
109	木曜日/寝室	People In The Box
110	蜃気楼の国	sora tob sakana
111	DESIRE	LUNA SEA
112	kurage	ぼくのりりっくのぼうよみ
113	youthful days	Mr.Children
114	ダメ、ゼツタイ、現実逃避	ReVision of Sence
115	なまけもの	サイダーガール
116	何も無い休日	キュウソネコカミ
117	生きてることが辛いなら	森山直太朗
118	部屋と地球儀	尾崎リノ
119	Kaonashi	androp
120	キエルツェの螺旋	the cabs
121	グレイの海	Eve
122	つじつま合わせに生まれた僕等	amazarashi
123	花の名	BUMP OF CHICKEN
124	割る!	岡崎体育+JINRO
125	「聴きたかったダンスミュージック、リキッドルームに」	サカナクション
126	独りんぼエンヴィー	majiko
127	龍の子	崎山蒼志
128	115万キロのフィルム	Official髭男dism
129	Pale Blue	米津玄師
130	インフェルノ	Mrs. GREEN APPLE
131	カタオモイ	Aimer
132	キンモクセイ	オレンジスパイニククラブ
133	トレモロ	RADWIMPS
134	ピーターパン	優里
135	ララバイ	RADWIMPS
136	わたがし	back number
137	愛を知るまでは	あいみょん
138	怪物	YOASOBI
139	勘弁えて悔しいわ	ずっと真夜中でいいのに。
140	瞳	大原櫻子
141	虹	菅田将暉
142	不思議	星野源
143	別の人の彼女になったよ	wacci
144	魔法の絨毯	川崎鷹也
145	又三郎	ヨルシカ
146	10月無口な君を忘れる	あたらしい
147	24H	SEVENTEEN
148	Film out	BTS
149	Re:Re:	ASIAN KUNG-FU GENERATION
150	THE OVER	UVERworld

表 3.4 実験で使用した楽曲の一覧(151~200 曲目)

音源番号	楽曲名	アーティスト名
151	Your eyes tell	BTS
152	ありがとう Forever	西内まりや
153	ある日、頭からツノが生えた	TOMORROW X TOGETHER
154	シンクロシティ	乃木坂46
155	ニルバナ	Tia
156	ブルーベリー・ナイツ	マカロニえんぴつ
157	狂	GEZAN
158	脱法ロック	Neru
159	六等星の夜	Aimer
160	檸檬	BURNOUT SYNDROMES
161	BUBBLE-GUM MAGIC	KEYTALK
162	CRAZY FOR YOUの季節	Base Ball Bear
163	drop pop candy feat. ギガ	Reol
164	オシャレ大作戦	ネクライトーキー
165	ハロ/ハワユ	CIVILIAN
166	リセット	向井太一
167	風神	人間椅子
168	未来	コブクロ
169	Grand Blue	湘南乃風
170	コーヒーとシロップ	Official髭男dism
171	シャリユージュ	*Luna
172	スタンドバイミー	the peggies
173	ツキミソウ	Novelbright
174	ドクハク	Gero
175	ヒカリへ	miwa
176	ポラリス	ヒトリエ
177	ルーマー	ポリスピカデリー
178	私へ ~光にまみれたver.~	GOOD ON THE REEL
179	手紙 ~拝啓十五の君へ~	アンジェラ・アキ
180	正解(18FES ver.)	RADWIMPS
181	誠 -Live for Justice-	浦島坂田船
182	放課後色	め組
183	きらり	藤井風
184	チェリボム	Silent Siren
185	ハローグッバイ	江口拓也
186	巡る巡る	神はサイコロを振らない
187	THE DAY	ボルノグラフィティ
188	us	milet
189	スターマーカー	KANA-BOON
190	スマイル	森七菜
191	バイバイサンキュー	BUMP OF CHICKEN
192	ピースサイン	米津玄師
193	炎	LiSA
194	時代	中島みゆき
195	第六感	Reol
196	未完成	家入レオ
197	Night Diver	三浦春馬
198	STACKiNG	BiSH
199	Teenager Forever	King Gnu
200	YOKAZE	変態紳士クラブ

3.1.4 アンケート

3.1.4.1 聴取後の評価

実験では聴取後の感情評価を取得するために図 3.2 に示すアンケートを使用した。このアンケートにより、聴いた各楽曲に対して「楽しい」、「悲しい」、「イライラする」、「落ち着く」の 4 種類の感情をそれぞれどの程度抱いたかを 5 段階の評価で取得した。

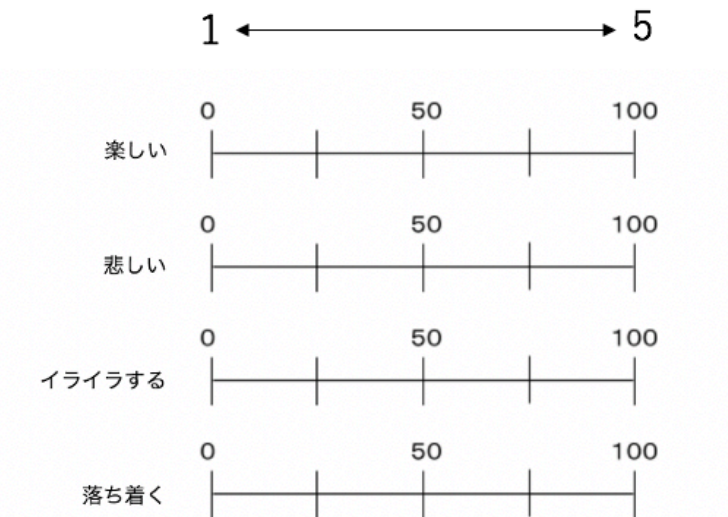


図 3.2 聴取後のアンケート

3.1.4.2 聴取中の評価

実験では聴取中の感情評価を取得するために DARMA[14]とジョイスティックを使用した。ここで DARMA とは映像や音楽を視聴しながら、ジョイスティック等によるユーザデバイスを使用して、2次元空間中の座標データを連続的に取得できるツールのことである。このツールを使用して、ラッセルの円環モデルの Arousal と Valence に関し、それぞれ-1~1の連続値で評価を取得した。聴取中の感情評価のアンケートの詳細を図 3.3 に、聴取中の感情評価取得の様子を図 3.4 に示す。

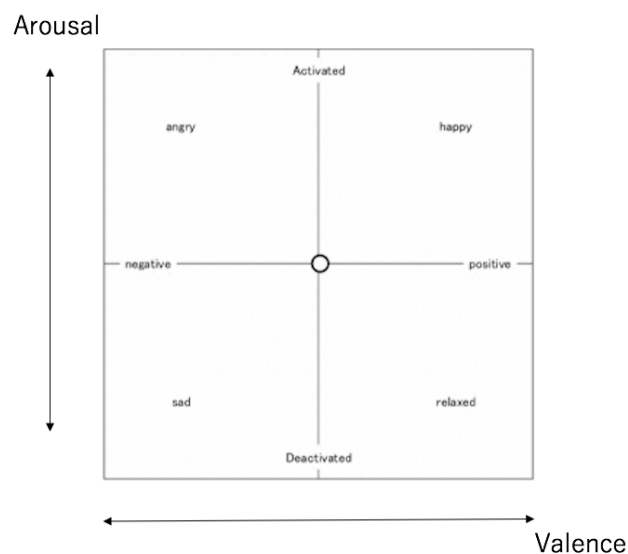


図 3.3 聴取中のアンケート

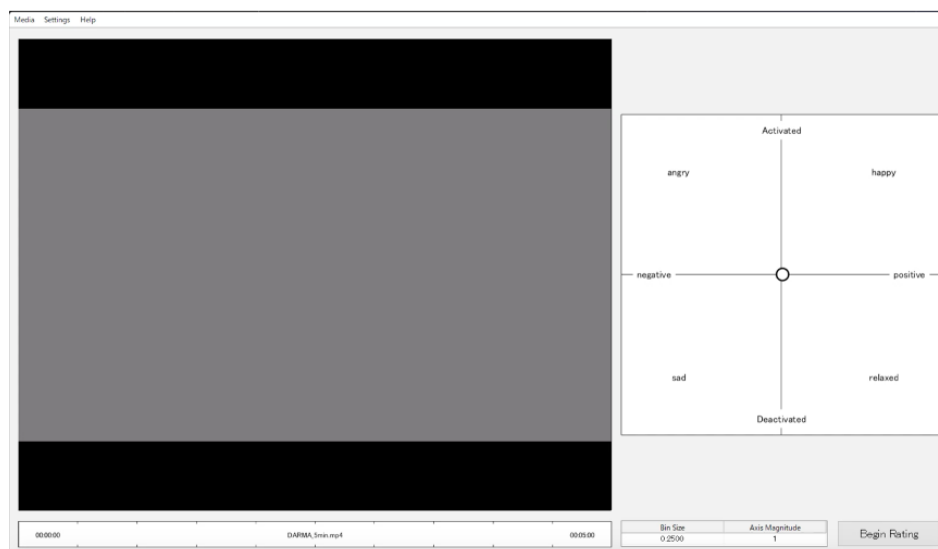


図 3.4 聴取中の感情評価取得の様子

3.2 実験手順

被験者は脳波計、心拍計、ノイズキャンセリングイヤホンを装着して実験を行った。また、被験者は瞳孔径測定のために、楽曲聴取中はパソコンに表示されているグレー画面の中心を見ていた。このことにより、対光反射補正を考慮しなくてよいとした。

まず、通常の状態の生体情報を得るために、無音のトラックを 60 秒間聴取した。その後、被験者は 1 曲ずつ楽曲を聴き、各曲の再生中と再生後にその楽曲に対する感情評価を行なった。5 回の試聴と評価が終わるたびに、被験者は短い休憩をとった。この実験の流れを図

にしたものを図 3.5 に示す. この流れを 1 セット分とし, 1 日に数セット行ない, 数日間にわたって実験を行なうことで 200 曲分のデータの収集を行なった.

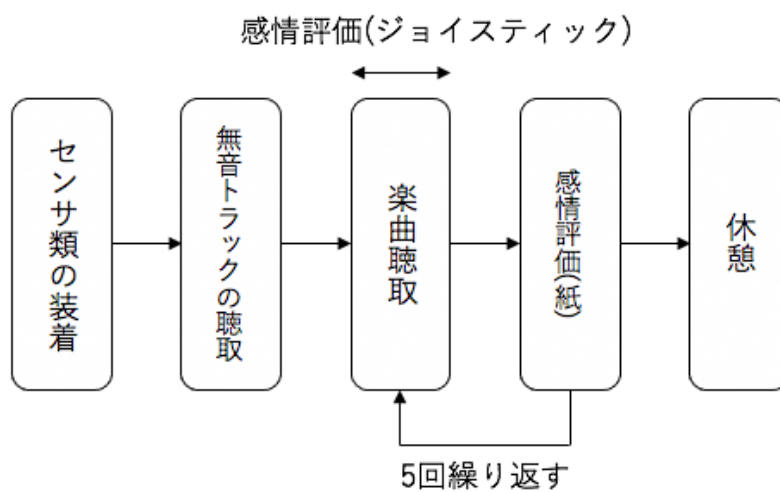


図 3.5 実験の流れ(1 セット分)

第4章 解析データと解析方法

はじめに、実験で取得した生体情報に関して、解析には基礎律動、RRI、瞳孔径の全てのデータにおいて最初と最後の1秒を除いたデータを用いた。また、聴取後の評価は表3.1~表3.4における音源番号1~200の200曲分のデータ、聴取中の評価は表3.1~表3.4における音源番号101~200の100曲分(平均再生時間253.61秒、標準偏差49.51)のデータとなっている。

4.1 分類による解析

4.1.1 聴取後と聴取中の評価を使用した場合の比較

楽曲内における感情の変化の考慮ありと考慮なしの比較として、聴取後の評価と聴取中の評価それぞれから正解ラベルを作成し、これらを目的変数、生体情報を説明変数として、解析を行い、推定精度の比較を行った。また、同時に内挿と外挿の両方の場合で検証を行ない、これらについても推定精度の比較を行なった。本検証では聴取後の評価から作成した正解ラベルを使用した場合は200曲分のデータ、聴取中の評価から作成した正解ラベルを使用した場合は100曲分のデータとなっている。

4.1.1.1 データ整形

基礎律動に関して、まず、瞬き等のノイズ除去のためにEEGLAB[15]とMARA[16]を使用して独立成分分析を行ない、14個の独立成分を想定したのち、ノイズを手動で選択し、除去した。200曲分のデータに関して、残った独立成分の平均の数は5.90であり、標準偏差は1.63である。その後、残った独立成分で再構成を行なった。1曲あたりのノイズ除去の様子の例を図4.1~図4.3に示す。次に、チャンネルごとに1秒窓、0.25秒スライドで高速フーリエ変換(FFT)を施し、1~45[Hz]の周波数スペクトルを算出し、各チャンネルで α 波(8-13[Hz])、 β 波(14-30[Hz])、 γ 波(31-45[Hz])に分けた。その後、図3.1から見て取れるように各チャンネルが頭の左右対称の位置にあることを利用して、 α 波、 β 波それぞれについて、左右比と左右差を算出した。また、それらとは別に、各チャンネルで β/α 、 γ/α を算出した。

RRIに関して、毎秒のRRIを無音トラック聴取時の平均RRIで割ることでRRI変化率を求め、その後0.25秒ごとに線形補間を行った。さらに、算出された値のうち150より大きいものを欠損区間の値と想定して除去した。

瞳孔径に関して、まず、瞬目の前後0.05秒は瞬きの途中であると想定し、瞬目の前後0.05秒も欠損として線形補間を行った。次に、左右の瞳孔径の平均を算出した。その後、1秒窓、0.25秒スライドで移動平均を算出した。

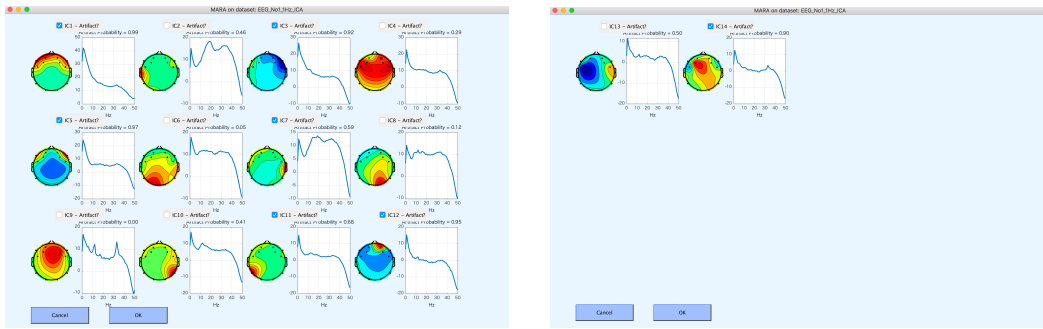


図 4.1 EEGLAB による独立成分分析の結果

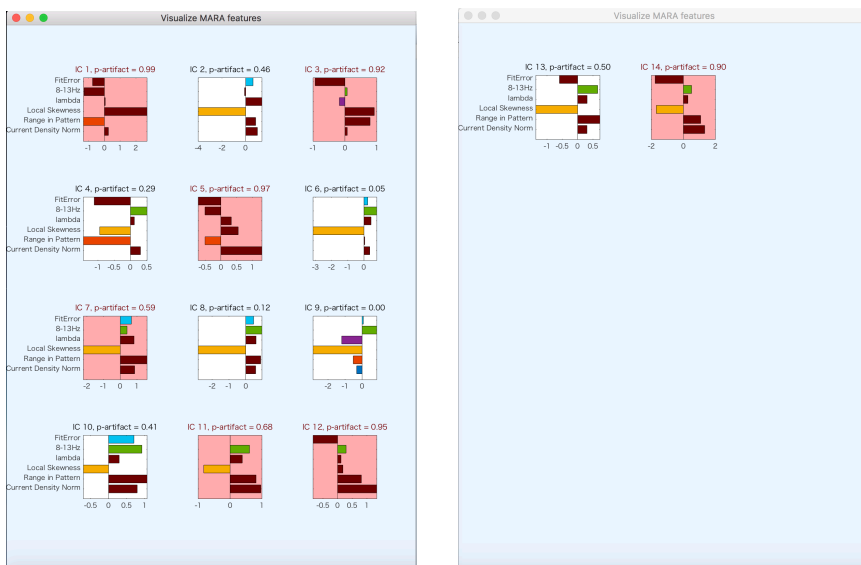


図 4.2 MARA によるノイズ成分の選択の様子

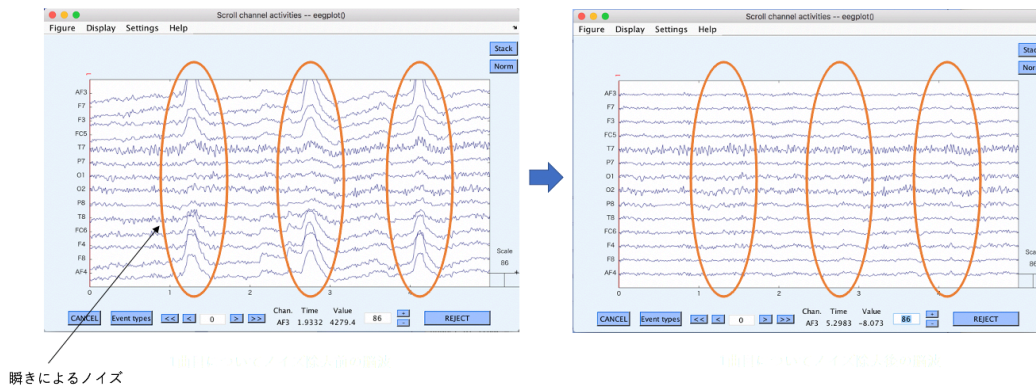


図 4.3 ノイズ除去前とノイズ除去後(再構成後の結果)の比較

4.1.1.2 正解ラベルの作成

聴取後の評価, 聴取中の評価それぞれから正解ラベルの作成を行なった.

聴取後の評価に関して, まず, 「悲しい」と「イライラする」についての回答はアンケート回答の偏りが大きかったため解析には使用せず, 「楽しい」と「落ち着く」の回答のみを使用した. 聴取後の評価からは 2 種類の方法で正解ラベルの作成を行った. 1 つ目として, アンケートの 5 段階評価(1~5)のうち, 3~5 の評価をその感情を強く抱いたとし, 1~2 の評価をその感情をそれほど抱かなかったとして感情ごとに 2 種類に分類するラベルを作成した. 2 つ目として, 1 つ目で感情ごとに作成したものを組みあわせて, 「楽しいかつ落ち着く」, 「楽しい」, 「落ち着く」, 「無感情」の 4 種類に分類するラベルを作成した. これらの聴取後の評価から作成したラベルは 1 楽曲につき 1 つの評価であるため, 各楽曲のデータ数に合うように値を繰り返して使用している.

聴取中の評価に関して, ラッセルの円環モデルの **Arousal** と **Valence** に関してそれぞれ 1~1 の連続値で取得しており, ここから聴取後の評価と比較が行ないやすいように, 「楽しい」に関しては **Arousal**, **Valence** ともに 0 以上でかつ座標の中心である(0,0)からの距離が 0.5 以上となる場合のみを強く楽しいと感じたとし, それ以外の場合をそれほど楽しいと感じなかったとして, 2 種類に分類するラベルに変換した. 「落ち着く」に関して **Arousal** が 0 以下かつ **Valence** は 0 以上でさらに座標の中心である(0,0)からの距離が 0.5 以上となる場合のみを強く落ち着くと感じたとし, それ以外の場合をそれほど落ち着くと感じなかったとして, 2 種類に分類するラベルに変換した. 聴取後の評価から作成した「楽しい」と「落ち着く」の 2 種類の感情の組み合わせによる分類のラベルと比較するためのラベルは聴取中の評価からは作成していない. また, 聴取中の評価は 20[Hz]で取得しているため, 上記の処理を各時刻で行なったあと, 1 秒窓, 0.25 秒スライドで最頻値を算出し, 取得した値を正解ラベルとした. ここで最頻値を算出した際に各評価の数が同数となった場合はそれほどその感情を抱かなかった評価として使用した.

4.1.1.3 解析手法

本検証では **Randomforest**, **DNN**, **CNN** の 3 種類の教師あり学習で解析を行なった. また, どの手法においても各楽曲内でランダムに学習データとテストデータを 8:2 に分ける内挿と楽曲単位で学習データとテストデータを 9:1 にわける外挿の 2 種類の方法で検証した. 外挿に関しては, 全楽曲が 1 回ずつテストデータとなるように交差検証も行なった.

Randomforest は機械学習アルゴリズムの 1 つであり, 決定木による複数の弱学習器を統合させて汎化能力を向上させるアンサンブル学習のアルゴリズムである [17]. **Randomforest** は **scikit-learn**[18]を用いて実装した. **Randomforest** の実行条件を表 4.1 に示す. 表 4.1 に示すパラメータ以外のパラメータの値はデフォルトの値となっている. また, 入力するデータセットは基礎律動(左右差, 左右比), **RRI**, 瞳孔径で構成されている. この入力するデータセットの構成は **DNN**, **CNN** と異なるものとなっているが, 前実験で精度がよ

かった組み合わせを入力データの構成としている。入力するデータセットは **z-score** により正規化している。

DNN(Deep Neural Network)は人間の脳神経系のニューロンを数理モデル化したものの組み合わせである **NN(Neural Network)**[19]をもとにし、ディープラーニングに対応させて4層以上に層を深くしたもので教師あり学習の1つである[20]。DNNは **Keras**[21]を使用して実装し、バックエンドには **Tensorflow**[22]を用いた。DNNの実行条件を表4.2に示す。また、入力するデータセットは基礎律動(左右比, β/α , γ/α), **RRI**, 瞳孔径で構成されている。この入力するデータセットの構成は **Randomforest** と異なるものとなっているが、**Randomforest** 同様、前実験で精度がよかった組み合わせを入力データの構成としている。入力するデータセットは **z-score** により正規化している。

CNN は一般的な順伝播型のニューラルネットワークとは異なる、全結合層の他に、畳み込み層(**Convolution Layer**)やプーリング層(**Pooling Layer**)から構成されるニューラルネットワークのことであり、教師あり学習の1つである[23]。CNNは **Keras**[21]を使用して実装し、バックエンドには **Tensorflow**[22]を用いた。CNNのネットワーク構成を図4.4に示す。次に、CNNのパラメータを表4.3に示す。また、CNNではネットワークの入力に向けて、学習データとテストデータに分けたのち、楽曲ごとに **1step** ずつずらしながら **58次元**×**32step** のデータを切り出すデータ整形を行っている。この作業を図にしたものを図4.5に示す。入力するデータセットに関しては、基礎律動(左右比, β/α , γ/α), **RRI**, 瞳孔径で構成されている。この入力するデータセットの構成は **Randomforest** と異なるものとなっているが、**Randomforest** や **DNN** 同様、前実験で精度がよかった組み合わせを入力データの構成としている。入力するデータセットは **z-score** により正規化している。

表 4.1 **Randomforest** の実行条件

min_samples_leaf	n_estimators
5	100

表 4.2 **DNN** の実行条件

層構造	損失関数		最適化関数
58→64→64→32→2 or 4	categorical_crossentropy		Adagrad
バッチサイズ	エポック数	活性化関数	学習率
128	300	ReLU	0.0005

表 4.3 CNN の実行条件

損失関数		最適化関数	
categorical_crossentropy		Adagrad	
バッチサイズ	エポック数	活性化関数	学習率
128	50	ReLU	0.0005

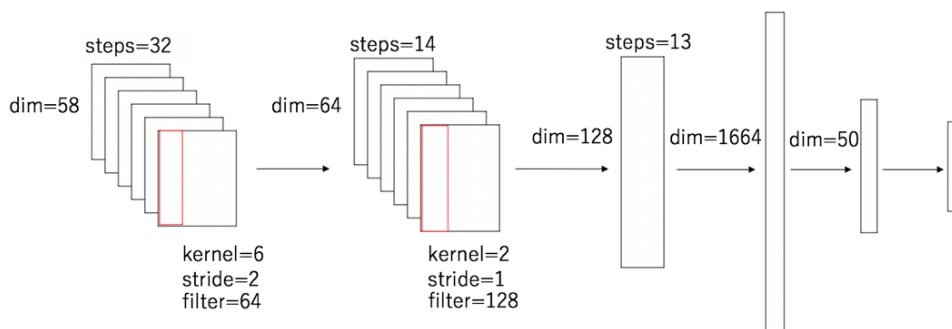


図 4.4 CNN のネットワーク構成

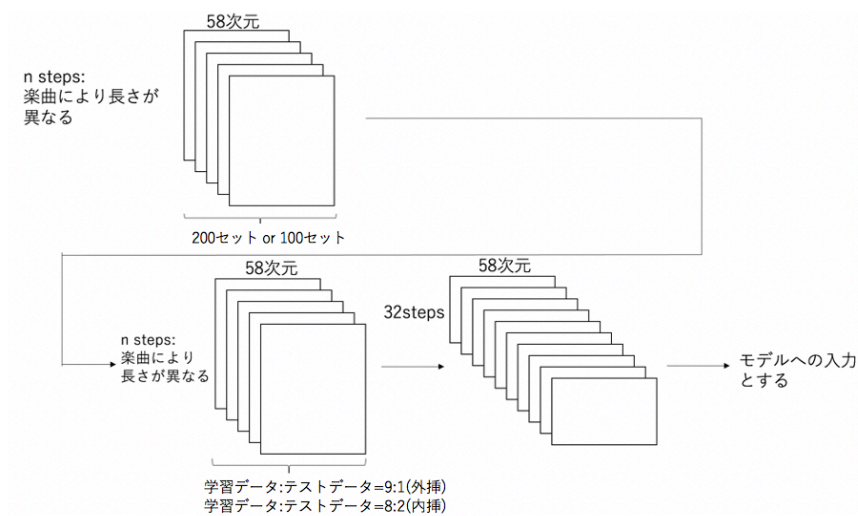


図 4.5 CNN のネットワークに入力する際のデータ整形

4.1.2 楽曲情報を追加した解析

Spotify[4]より SpotifyAPI[24]を使用して、各楽曲に関する 4 種類の情報(loudness, tempo, key, mode)を取得し、解析に使用した。実験に使用した楽曲のうち、Spotify で配信されていない楽曲が一部存在していたため本検証で解析に使用したデータは 181 曲分となっている。生体情報のみの場合、楽曲情報のみの場合、生体情報と楽曲情報の場合の 3 種類の入力データセットを作成し、推定精度の比較を行なった。

4.1.2.1 データ整形

生体情報(基礎律動, RRI, 瞳孔径)に関して, データ整形の方法は 4.1.1.1 節にて使用した方法と同様の手順でデータ整形を行なった.

Spotify[4]より取得した, 各楽曲情報に関しては 1 楽曲につき 1 つの値として得られるので各楽曲のデータ数に合うように値を繰り返し, 各楽曲の長さ分にして使用した.

4.1.2.2 正解ラベルの作成

本検証では聴取後の評価のみを使用して正解ラベルの作成を行なった. 聴取後の評価から 4.1.1.2 節で作成した正解ラベルと同様の手順で感情ごとに 2 種類に分類するラベルと感情の組み合わせにより 4 種類に分類するラベルを作成した.

4.1.2.3 解析手法

DNN や CNN を用いた場合, 楽曲情報のみのデータセットにおいて, データのバリエーションが少なく, 学習が上手く進まなかったため本検証では **Randomforest** のみを使用して解析を行なった. また, 楽曲単位で学習データとテストデータを 9:1 にわける外挿で検証を行なった. 全楽曲が 1 回ずつテストデータとなるように交差検証も行った. **Randomforest** の実行条件は 4.1.1.3 節と同様であり, 表 4.1 に示すパラメータ以外のパラメータの値はデフォルトの値となっている. また, 入力するデータセットは生体情報のみの場合は基礎律動(左右差, 左右比), RRI, 瞳孔径, 楽曲情報のみの場合は楽曲情報(loudness, tempo, key, mode), 生体情報と楽曲情報の場合は基礎律動(左右差, 左右比), RRI, 瞳孔径, 楽曲情報(loudness, tempo, key, mode)で構成されている. 入力するデータセットは z-score により正規化している.

4.1.3 AutoEncoder によるデータ拡張を使用した解析

外挿での推定精度向上に向けて **AutoEncoder** により生体情報のデータ拡張を行ない, 取得した拡張データを追加して解析を行なった. 本検証では使用したデータは 200 曲分となっている.

4.1.3.1 データ整形

生体情報(基礎律動, RRI, 瞳孔径)のみを使用した. データ整形の方法に関しては, 4.1.1.1 節にて使用した方法と同様の手順でデータ整形を行なった.

4.1.3.2 正解ラベルの作成

本検証では聴取後の評価のみを使用して正解ラベルの作成を行なった. 聴取後の評価から 4.1.1.2 節で作成した正解ラベルと同様の手順で感情ごとに 2 種類に分類するラベルと感情の組み合わせにより 4 種類に分類するラベルを作成した.

4.1.1.3 解析手法

本検証ではまず **AutoEncoder** によりデータ拡張を行い、その後 **DNN** により感情推定の解析を行なった。ここで、**DNN** を解析に使用したのは 4.1.1 節にて **DNN** を使用した場合に推定精度が一番高くなったためである。詳細な解析の流れは次の通りである。まず、200 曲分のデータから楽曲単位で学習データとテストデータを 9:1 に分け、このテストデータを **DNN** 用のテストデータとした。残りの学習データから **AutoEncoder** 用に学習データとテストデータを各楽曲内でランダムに 8:2 に分け、これらを使用して **AutoEncoder** を学習させた。**AutoEncoder** の学習後、**AutoEncoder** 用の学習データ、テストデータの両方を学習済み **AutoEncoder** に入力し、デコーダから拡張データを取得した。この取得した拡張データと学習済み **AutoEncoder** に入力したデータの両方を **DNN** 用の学習データとした。用意した **DNN** 用の学習データとテストデータを用いて感情推定のための **DNN** の学習と検証を行なった。以上の解析の流れにおける学習データとテストデータの分割方法を図にしたものを図 4.6 に示す。また **DNN** における検証に関して、交差検証も行なった。

AutoEncoder はニューラルネットワークの仕組みの一つであり、入力されたデータをエンコーダにより一度圧縮し、重要な特徴量だけを残したあと、デコーダにより、もとの次元に復元処理をするアルゴリズムである。このエンコーダより出力されるデータは入力データを重要な特徴を含みつつ、より少ない次元で表現したものと考えられるため、次元圧縮手法としてよく用いられる。他にもデコーダより出力されるデータは残った重要な特徴量のみから復元されたデータとなるため、データクレンジングの手法としても用いられることがある[25][26]。**AutoEncoder** の概念図を図 4.7 に示す。**AutoEncoder** は **Keras**[21] を使用して実装し、バックエンドには **Tensorflow**[22] を用いた。本検証で使用した **AutoEncoder** のネットワーク構成を図 4.8 に、**AutoEncoder** の実行条件を表 4.4 に示す。また、入力するデータセットは基礎律動(左右比、 β/α 、 γ/α)、**RRI**、瞳孔径で構成されており、 z -score により正規化している。

本検証で使用した **DNN** の実行条件は 4.1.1.3 節と同様であり、表 4.2 の通りである。**AutoEncoder** 同様、入力するデータセットは基礎律動(左右比、 β/α 、 γ/α)、**RRI**、瞳孔径で構成されている。

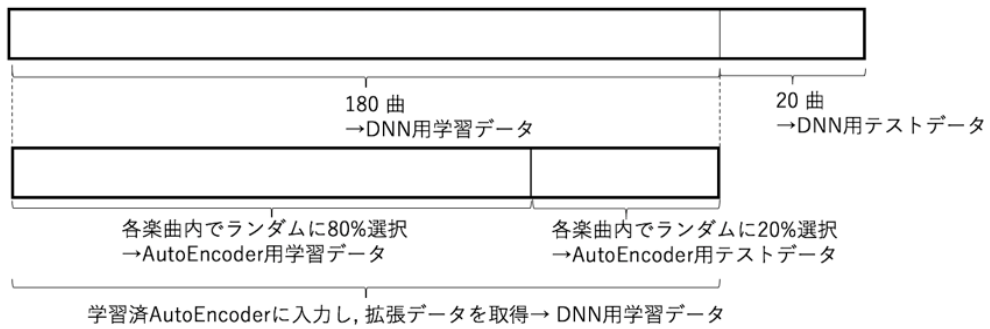


図 4.6 学習データとテストデータの分割方法

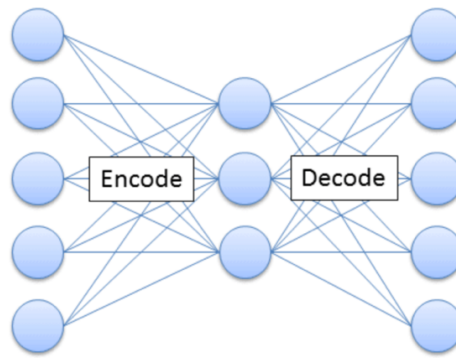


図 4.7 AutoEncoder の概念図(文献[25]の図より引用)

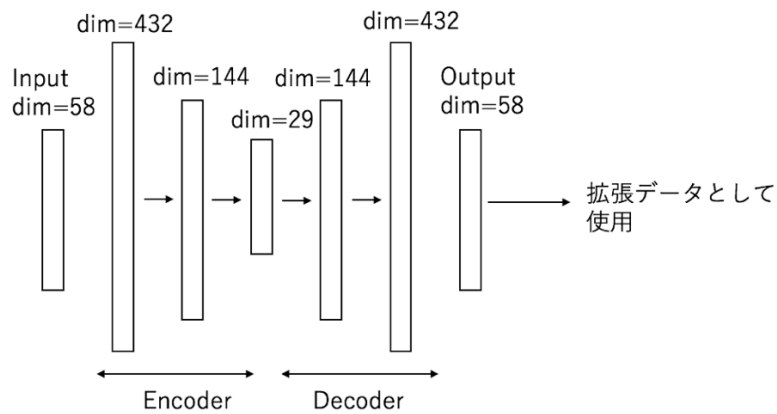


図 4.8 AutoEncoder のネットワーク構成

表 4.4 AutoEncoder の実行条件

損失関数		最適化関数	
MSLE		Adam	
バッチサイズ	エポック数	活性化関数	学習率
128	250	tanh	0.001

4.2 回帰による解析

楽曲内における感情の変化を考慮した解析として、4.1.1 節のように取得した聴取中の評価の値をクラスに変換せず、数値として利用する解析も行なった。聴取中の評価をラッセルの円環モデルの **Arousal** と **Valence** に関してそれぞれ-1~1 の連続値で取得していることを利用して、回帰での解析を行ない、得られた結果を決定係数(**R2score**)と平均平方二乗誤差(**RMSE**)により評価した。

4.2.1 データ整形

生体情報(基礎律動, **RRI**, 瞳孔径)のみを使用した。データ整形の方法に関しては、4.1.1.1 節にて使用した方法と同様の手順でデータ整形を行なった。

4.2.2 正解ラベルの作成

聴取中の評価はラッセルの円環モデルの **Arousal** と **Valence** に関してそれぞれ-1~1 の連続値(20[Hz]計測)で取得している。ここから **Arousal** と **Valence** それぞれについて、1 秒窓、0.25 秒スライドで移動平均を算出し、これを正解ラベルとした。

4.2.3 解析手法

本検証では勾配ブースティング決定木(**GBDT**)を使用して解析を行なった。各楽曲内でランダムに学習データとテストデータを 8:2 に分ける内挿と楽曲単位で学習データとテストデータを 9:1 にわける外挿の 2 種類の方法で検証した。外挿に関しては、全楽曲が 1 回ずつテストデータとなるように交差検証も行なった。

勾配ブースティング決定木は、一つ前の決定木の誤差を次の決定木が修正するように繰り返しながら、決定木を追加していくもので、アンサンブル学習の一つである[27]。勾配ブースティング決定木は **LightGBM**[28]を使用して実装した。**GBDT** の実行条件について **Arousal** のものを表 4.5 に、**Valence** のものを表 4.6 に示す。実行条件が **Arousal** と **Valence** で異なるのはどちらかに合わせた場合、学習が上手く進まなかったためであり、それぞれの場合で **Optuna**[29]を使用してパラメータチューニングを行なった結果を示している。また表 4.5 と表 4.6 に示すパラメータ以外のパラメータの値はデフォルトの値となっている。入力するデータセットは基礎律動(左右差, 左右比), **RRI**, 瞳孔径で構成されており、**z-score** により正規化している。

表 4.5 **GBDT** の実行条件(**Arousal**)

boosting_type	metric	learning_rate	num_leaves	min_data_in_leaf	max_depth	lambda_l2
GBDT	RMSE	1.5	10	5	10	9.42606
lambda_l1	feature_fraction	bagging_fraction	bagging_freq	num_iteration	min_child_samples	
9.40769	1.00000	0.95540	4	100	10	

表 4.6 GBDT の実行条件(Valence)

boosting_type	metric	learning_rate	num_leaves	min_data_in_leaf	max_depth	lambda_l2
GBDT	RMSE	1.5	6	5	10	7.54395E-07
lambda_l1	feature_fraction	bagging_fraction	bagging_freq	num_iteration	min_child_samples	
5.28102	0.75200	1.00000	0	100	25	

第5章 結果と考察

5.1 分類による解析

5.1.1 聴取後と聴取中の評価を使用した場合の比較

内挿による検証について得られた結果を表 5.1 に、外挿による検証について得られた結果について、聴取後の評価を使用した場合のものを表 5.2 に、聴取中の評価を使用した場合のものを表 5.3 に示す。

表 5.1 より内挿による検証に関してはどの場合においても高い精度で推定ができていことがわかる。一方で表 5.2 と表 5.3 より外挿による検証ではどの場合においても交差検証で取得した平均の推定精度はチャンスレベルを超える精度とはなっているものの、内挿ほど高い精度で推定できていないことがわかる。この外挿による検証の結果について、どの場合でも平均の推定精度がチャンスレベルを超える精度になったことから、なんらかの特徴は捉えることができていると考える。一方で、内挿による検証に比べ、外挿による検証において推定精度が低くなっていることから、聴取時の感情分類に利用できるような特徴をまだそれほど捉えられていないと推測する。そしてこのような結果になった原因として、実験を行なった日の状態などによるノイズがうまく前処理で除くことができていないことやデータ数がまだ少ないことなどが考えられる。

次に表 5.2 と表 5.3 の「楽しい」や「落ち着く」の結果に関して、聴取中の評価を使用した方が聴取後の評価を使用した場合より高い精度で推定できているように見える。しかし、ラベルの偏りが表 5.4 と表 5.5 に示すように聴取後の評価では 1:1 程度であるのに対し、聴取中の評価では 4:1 程度となっているため、片方のラベルのみに予測する可能性を考慮すると聴取後の評価を使用する方が推定に向いているという結果になっていると考える。この結果について、聴取中の評価を使用した解析の方が聴取後の評価を使用した解析に比べ、データ数が少ない上にラベルの偏りが大きいことが一番の原因だと考える。また、聴取中の評価について、現在の取得方法では同時に複数感情を抱く可能性があることを考慮できていないことや感情が生体情報にあらわれるまでの時間と感情を認識し、ジョイスティックを動かすまでの時間に差がある可能性を考慮できていないことなどもこのような結果になった原因の一つだと推測する。

表 5.1 内挿による検証の結果(推定精度)

	聴取後の評価			聴取中の評価	
	4分類	楽しい	落ち着く	Arousal	Valence
Randomforest	0.950	0.960	0.965	0.895	0.867
DNN	0.673	0.804	0.776	0.844	0.811
CNN	0.990	0.998	0.999	0.930	0.893

表 5.2 聴取後の評価を使用した場合の外挿による検証の結果(推定精度)

Fold #	Randomforest			DNN			CNN		
	4分類	楽しい	落ち着く	4分類	楽しい	落ち着く	4分類	楽しい	落ち着く
Fold 1	0.287	0.464	0.506	0.356	0.544	0.533	0.363	0.430	0.463
Fold 2	0.294	0.622	0.471	0.310	0.691	0.409	0.336	0.717	0.403
Fold 3	0.384	0.580	0.500	0.312	0.623	0.511	0.278	0.580	0.525
Fold 4	0.336	0.575	0.540	0.365	0.637	0.557	0.401	0.667	0.469
Fold 5	0.409	0.607	0.530	0.320	0.626	0.563	0.332	0.727	0.520
Fold 6	0.370	0.625	0.539	0.357	0.619	0.513	0.385	0.638	0.565
Fold 7	0.340	0.584	0.563	0.376	0.569	0.552	0.281	0.522	0.609
Fold 8	0.281	0.556	0.597	0.351	0.546	0.617	0.301	0.626	0.642
Fold 9	0.211	0.477	0.468	0.292	0.612	0.505	0.189	0.548	0.491
Fold 10	0.279	0.590	0.478	0.383	0.662	0.593	0.375	0.617	0.503
Average	0.319	0.568	0.519	0.343	0.613	0.536	0.324	0.607	0.519

表 5.3 聴取中の評価を使用した場合の外挿による検証の結果(推定精度)

Fold #	Randomforest		DNN		CNN	
	楽しい	落ち着く	楽しい	落ち着く	楽しい	落ち着く
Fold 1	0.814	0.747	0.871	0.755	0.675	0.661
Fold 2	0.778	0.799	0.790	0.853	0.533	0.709
Fold 3	0.853	0.765	0.854	0.787	0.741	0.742
Fold 4	0.815	0.700	0.818	0.717	0.770	0.703
Fold 5	0.786	0.764	0.811	0.766	0.691	0.764
Fold 6	0.830	0.785	0.830	0.811	0.750	0.761
Fold 7	0.826	0.826	0.841	0.830	0.750	0.731
Fold 8	0.819	0.782	0.819	0.837	0.649	0.640
Fold 9	0.703	0.741	0.701	0.746	0.706	0.752
Fold 10	0.771	0.749	0.774	0.759	0.660	0.600
Average	0.800	0.766	0.811	0.786	0.692	0.706

表 5.4 各ラベルの割合(聴取後の評価)

4分類				楽しい		落ち着く	
楽しいかつ落ち着く	楽しい	落ち着く	無感情	特に楽しい	それほど楽しくない	特に落ち着く	それほど落ち着かない
16.0%	28.6%	31.8%	23.6%	44.7%	55.3%	47.8%	52.2%

表 5.5 各ラベルの割合(聴取中の評価)

楽しい		落ち着く	
特に楽しい	それほど楽しくない	落ち着く	それほど落ち着かない
18.5%	81.5%	21.5%	78.5%

5.1.2 楽曲情報を追加した解析

用意した各データセットについて解析した結果を表 5.6 に示す.

表 5.6 より, どのラベルづけ方法においても楽曲情報を追加したことで推定精度が大きく向上し, 楽曲情報と生体情報の両方を用いた場合で推定精度が一番高くなることがわかる. このことから, 楽曲情報だけでは不十分であり, 生体情報を加えることで, より個人の嗜好を捉えられるようになる可能性があると考え. この結果は被験者 1 人分の結果であるため, より確実なものにするために他の被験者のデータについても解析する必要があると考え. また, 本検証では Spotify[4]より取得した loudness, tempo, key, mode の 4 種類の指標のみを解析に使用したが, 他にも楽曲自体の解析から得られるコード進行などの特徴や楽曲についての歌詞を解析して得られる歌詞の内容に関する特徴なども生体情報と合わせて解析に使用することでさらに推定精度を向上させることができる可能性があると考え.

表 5.6 楽曲情報を使用した解析の結果(推定精度)

Fold #	4分類			楽しい			落ち着く		
	生体情報+ 楽曲情報	楽曲情報 のみ	生体情報 のみ	生体情報+ 楽曲情報	楽曲情報 のみ	生体情報 のみ	生体情報+ 楽曲情報	楽曲情報 のみ	生体情報 のみ
Fold 1	0.336	0.324	0.314	0.700	0.623	0.610	0.626	0.475	0.465
Fold 2	0.400	0.371	0.254	0.808	0.858	0.521	0.539	0.453	0.613
Fold 3	0.510	0.470	0.387	0.757	0.760	0.741	0.662	0.499	0.539
Fold 4	0.392	0.457	0.309	0.749	0.733	0.528	0.493	0.324	0.540
Fold 5	0.368	0.469	0.333	0.703	0.700	0.615	0.516	0.446	0.613
Fold 6	0.513	0.379	0.406	0.766	0.613	0.460	0.489	0.591	0.543
Fold 7	0.409	0.263	0.386	0.580	0.595	0.555	0.417	0.537	0.433
Fold 8	0.555	0.340	0.340	0.822	0.777	0.542	0.488	0.636	0.490
Fold 9	0.295	0.233	0.258	0.515	0.470	0.513	0.883	0.836	0.583
Fold 10	0.297	0.416	0.254	0.688	0.641	0.662	0.486	0.458	0.458
Average	0.407	0.372	0.324	0.709	0.677	0.575	0.560	0.526	0.528

5.1.3 AutoEncoder によるデータ拡張を使用した解析

拡張データを追加した解析の結果を表 5.7 に示す.

表 5.7 と表 5.2 における DNN を使用した場合の平均推定精度を比較すると, 感情の組み合わせによる分類の場合では推定精度が向上した. 一方で, 感情ごとに 2 分類の場合は推定精度が低下する結果となっていることがわかる. また交差検証における Fold ごとの結果を詳細にみても推定精度が向上した Fold も推定精度が低下した Fold もあるということがわかる. 以上の結果から本検証で追加した拡張データが推定精度向上に対し, それほど効果的な影響を与えていないと考える. これは 5.1.1 節でも述べたことと同様に, データ数が少

なく, AutoEncoder でまだ上手く感情推定に関わる特徴を捉えられていないことや拡張前の入力データに実験を行なった日の状態などによるノイズが残っており, それをもとにした追加データの生成になっていることによると考える. また, 各拡張データが各感情の特徴を持ったデータになっているのかについて確認を行なっていないので, 拡張データに正しいラベルがついているのか定かではない. そのため, このことも上記のような結果になった原因の一つとなっている可能性があると考え.

表 5.7 拡張データを追加した解析の結果(推定精度)

Fold #	4分類	楽しい	落ち着く
Fold 1	0.331	0.568	0.571
Fold 2	0.351	0.651	0.463
Fold 3	0.343	0.570	0.569
Fold 4	0.463	0.613	0.540
Fold 5	0.441	0.597	0.507
Fold 6	0.383	0.605	0.444
Fold 7	0.337	0.572	0.539
Fold 8	0.343	0.569	0.665
Fold 9	0.298	0.533	0.479
Fold 10	0.321	0.617	0.536
Average	0.361	0.590	0.531

5.2 回帰による解析

内挿による検証の結果のうち Arousal について, テストデータに対する推定結果を図 5.1 に, Valence について, テストデータに対する推定結果を図 5.2 に示す. また, 内挿による検証の結果を R2score と RMSE で評価した結果を表 5.8 に示す. 同様に外挿による検証の結果のうち, Arousal について, テストデータに対する推定結果を図 5.3 に, Valence について, テストデータに対する推定結果を図 5.4 に示し, これらの結果を R2score と RMSE で評価した結果を表 5.9 に示す. ここで外挿による検証におけるテストデータに対する推定結果については R2score が一番大きくかつ RMSE が一番小さくなっている Fold についての結果のみを示す. また, 図 5.1~図 5.4 において, 青色の線で示されているものが実測値, 赤色の線で示されているものが予測値となっており, 横軸が Index, 縦軸が評価の値となっている.

まず, 図 5.1 と図 5.2 より, 内挿による検証に関しては Arousal と Valence とともに細かい評価の変化は推定できていないが, 大まかな評価の変化は推定することができていることがわかる. 一方で, 図 5.3 と図 5.4 をみると, 外挿による検証に関しては, 一番結果の良い Fold で Arousal と Valence とともにわずかに大まかな評価の変化を推定できている程度とな

っており、他の Fold も含めると、ほとんど推定できていない結果となっている。また、表 5.8 と表 5.9 から、内挿による検証の結果では R^2 score が 0.4~0.5 程度の値となっており、RMSE が 0.3 程度の値となっているのに対し、外挿による検証の結果ではほとんどの Fold で R^2 score が負の値となっており、RMSE が 0.4~0.5 程度の値となっていることがみてとれる。モデルのあてはまりがよいほど R^2 score は大きく、実測値と予測値に近いほど RMSE は小さくなるため、上記の結果を踏まえると R^2 score と RMSE から外挿による検証では推定があまりできていないことがわかる。さらに、評価の指標が異なるため比較が難しいが、分類による解析と比較すると、分類による解析では外挿による検証においても推定精度から少しは推定できているようにみえるため、分類による解析の方が上手く推定ができていると考える。このようになった原因として分類による解析時同様、細かい感情の変化に関する特徴を捉えられるほどのデータ量がなく、データ量が足りていないことが考えられる。同時にこれが原因で回帰による解析が分類による解析ほど上手く推定できていない可能性もあると考える。他にも同時に複数感情を抱く可能性があることを考慮できていないことや感情が生体情報にあらわれるまでの時間と感情を認識し、ジョイスティックを動かすまでの時間に差がある可能性を考慮できていないことなども回帰による解析の方が上手く推定できていない原因となっている可能性がある。さらに内挿による検証ほど外挿による検証で推定ができていないことに関して、こちらも分類による解析時同様、実験時の状態によるノイズの影響が原因として考えられる。

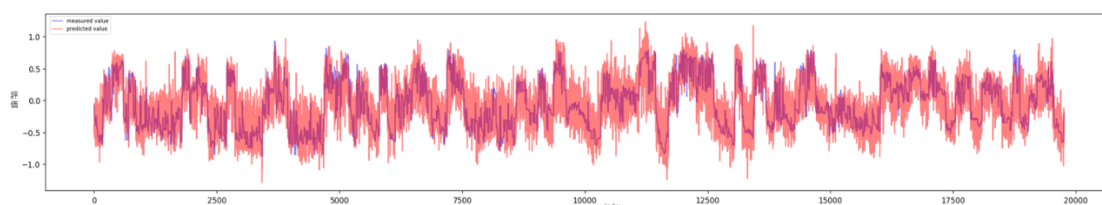


図 5.1 内挿についてテストデータに対する推定結果(Arousal)

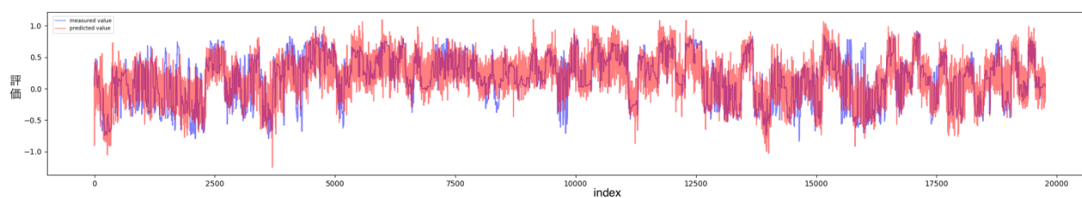


図 5.2 内挿についてテストデータに対する推定結果(Valence)

表 5.8 内挿による検証の結果を R2score と RMSE で評価した結果

	R2score	RMSE
Arousal	0.506	0.274
Valence	0.415	0.294

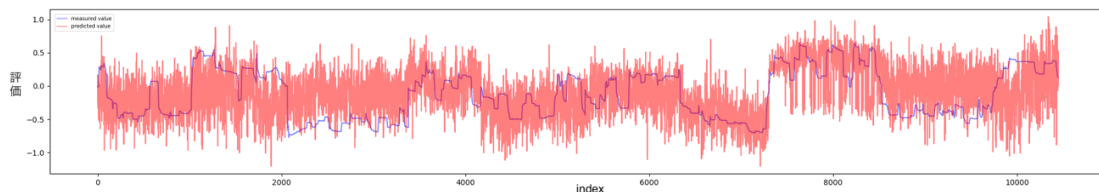


図 5.3 外挿について Fold 4 のテストデータに対する推定結果(Arousal)

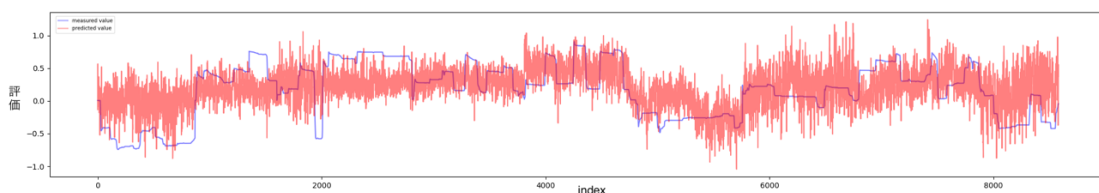


図 5.4 外挿について Fold 1 のテストデータに対する推定結果(Valence)

表 5.9 外挿による検証の結果を R2score と RMSE で評価した結果

Fold #	Arousal		Valence	
	R2score	RMSE	R2score	RMSE
Fold 1	-0.414	0.567	0.101	0.396
Fold 2	-0.959	0.555	-0.257	0.421
Fold 3	-0.433	0.421	-0.186	0.453
Fold 4	-0.261	0.409	-0.133	0.374
Fold 5	-0.538	0.510	-0.666	0.524
Fold 6	-0.729	0.519	-0.299	0.408
Fold 7	-0.480	0.467	-0.598	0.432
Fold 8	-0.950	0.454	-0.564	0.445
Fold 9	-0.911	0.477	-0.408	0.425
Fold10	-0.639	0.459	-0.276	0.433

第6章 結論

6.1 まとめ

本研究ではまず、取得データの量を増やすために被験者を筆者自身とした新実験を行い、外挿による検証を行なった。同時に、実験では楽曲聴取中もジョイスティックを用いて評価を取得し、これを使用して楽曲内での感情の変化を考慮した解析を行ない、楽曲内での感情の変化を考慮しなかった場合との比較を行なった。さらに、外挿による検証での推定精度向上のために Spotify[4]より取得した楽曲情報を利用した検証や AutoEncoder を用いたデータ拡張を追加した検証も行なった。本研究で得られた結果とそれに対する考察のまとめは次の通りである。

まず、聴取後の評価と聴取中の評価それぞれから正解ラベルを作成し、これらを目的変数、生体情報を説明変数として、解析を行い、推定精度の比較を行なった。また、同時に内挿と外挿の両方の場合で検証を行ない、これらについても推定精度の比較を行なった。その結果、内挿による検証と外挿による検証の比較に関しては、内挿による検証ではどの場合においても推定精度が高くなるが、一方で外挿による検証ではチャンスレベルを超える精度となっているものの、内挿ほど推定精度が高くないことがわかった。このことから、なんらかの特徴を捉えることができ、生体情報から感情の推定ができる可能性はあるが、一方で、ノイズの除去や解析に使用するデータ量に課題がまだ残されていると考える。解析に聴取後の評価を使用した場合と聴取中の評価を使用した場合の比較に関して、推定精度は聴取中の評価を使用した場合の方が高くなっているが、ラベルの偏りを考慮すると聴取後の評価を使用する方が推定に向いているという結果になっていると推測した。このことから解析に使用するデータ量に課題がまだあると考える。さらに聴取中の評価の取得方法や聴取中の評価からの正解ラベルの作成方法に関して課題が残されていると考える。

次に、Spotify[4]より取得した楽曲情報を利用し、生体情報のみの場合、楽曲情報のみの場合、生体情報と楽曲情報の場合の3種類の入力データセットを作成して推定精度の比較を行なった。その結果、楽曲情報と生体情報の両方を用いた場合で推定精度が一番高くなり、このことから楽曲情報だけでは不十分であり、生体情報を加えることで、より個人の嗜好を捉えられるようになる可能性が示唆された。

さらに、AutoEncoderによりデータ拡張を行い、その後DNNにより感情推定の解析を行なった。その結果、拡張データを追加しても推定精度の向上がみられる場合とみられない場合が存在し、全体をまとめてみるとデータ拡張による効果はそれほど得られなかった。一方で効果がそれほど得られなかった原因として入力データに残されたノイズによる影響や拡張データにつけられた正解ラベルが正しいかわからないことによる影響などが考えられるため、これらの課題を解決すればデータ拡張が有効となる可能性もあると考える。

最後に、聴取中の評価の値をクラスに変換せず、数値として利用し、回帰による解析を行

なった。その結果、内挿による検証では **Arousal** と **Valence** ともに細かい評価の変化の推定はできていないが、大まかな評価の変化の推定はできることがわかった。一方で、外挿による検証では交差検証の結果全体を通してみるとほとんど推定できていない結果となった。また、上記で述べた分類による解析の結果と比較すると、分類による解析の結果では外挿による検証においても推定精度から少しは推定できているように見えるため、分類による解析の方が上手く推定ができていると考える。このような結果になったのはデータ数の影響が大きいと考えるが、他にも同時に複数感情を抱く可能性があることを考慮できていないことや感情が生体情報にあらわれるまでの時間と感情を認識し、ジョイスティックを動かすまでの時間に差がある可能性を考慮できていないことが影響していることも考えられるためこれらの課題を解決すれば結果が変わる可能性があるかと推測する。

6.2 今後の課題

6.1 節を踏まえると全体を通して今後の課題としてまず、解析に使用するデータのさらなる追加とデータの前処理方法についての再検討が挙げられる。前処理方法についての再検討の例としては、実験を行なった日の状態によるノイズの除去方法の検討などがある。また、前処理方法の検討のみでなく、よりノイズの少ないデータを取得するために生体情報の取得方法について再検討することも有効である可能性があると考え。次に、本研究で使用したデータは被験者 1 人分のみであるため、他の被験者の場合でも同様の結果になるのかを確認する必要があると考える。他にも本研究で行なった各検証に注目すると、聴取中の評価の取得方法と聴取中の評価を使用した正解ラベルの作成方法の再検討、さらなる楽曲情報の利用、拡張データにつけられた正解ラベルが正しいかということを確認する方法の検討なども今後の課題として挙げられる。

図一覧

図 2.1 心電図における基本波形(文献[8]の図 2 より引用)	3
図 3.1 EPOCX の電極の位置(文献[13]の図より引用)	5
図 3.2 聴取後のアンケート	10
図 3.3 聴取中のアンケート	11
図 3.4 聴取中の感情評価取得の様子	11
図 3.5 実験の流れ(1 セット分)	12
図 4.1 EEGLAB による独立成分分析の結果	14
図 4.2 MARA によるノイズ成分の選択の様子	14
図 4.3 ノイズ除去前とノイズ除去後(再構成後の結果)の比較	14
図 4.4 CNN のネットワーク構成	17
図 4.5 CNN のネットワークに入力する際のデータ整形	17
図 4.6 学習データとテストデータの分割方法	20
図 4.7 AutoEncoder の概念図(文献[25]の図より引用)	20
図 4.8 AutoEncoder のネットワーク構成	20
図 5.1 内挿についてテストデータに対する推定結果(Arousal)	27
図 5.2 内挿についてテストデータに対する推定結果(Valence)	27
図 5.3 外挿について Fold 4 のテストデータに対する推定結果(Arousal)	28
図 5.4 外挿について Fold 1 のテストデータに対する推定結果(Valence)	28

表一覧

表 3.1 実験で使⽤した楽曲の一覧(1~50 曲目)	6
表 3.2 実験で使⽤した楽曲の一覧(51~100 曲目)	7
表 3.3 実験で使⽤した楽曲の一覧(101~150 曲目)	8
表 3.4 実験で使⽤した楽曲の一覧(151~200 曲目)	9
表 4.1 Randomforest の実行条件	16
表 4.2 DNN の実行条件	16
表 4.3 CNN の実行条件	17
表 4.4 AutoEncoder の実行条件	20
表 4.5 GBDT の実行条件(Arousal)	21
表 4.6 GBDT の実行条件(Valence)	22
表 5.1 内挿による検証の結果(推定精度)	23
表 5.2 聴取後の評価を⽤した⼾の外挿による検証の結果(推定精度)	24
表 5.3 聴取中の評価を⽤した⼾の外挿による検証の結果(推定精度)	24
表 5.4 各ラベルの割合(聴取後の評価)	24
表 5.5 各ラベルの割合(聴取中の評価)	24
表 5.6 楽曲情報を⽤した解析の結果(推定精度)	25
表 5.7 拡張データを追加した解析の結果(推定精度)	26
表 5.8 内挿による検証の結果を R2score と RMSE で評価した結果	28
表 5.9 外挿による検証の結果を R2score と RMSE で評価した結果	28

付録

付録 A 回帰による解析におけるテストデータに対する推定結果

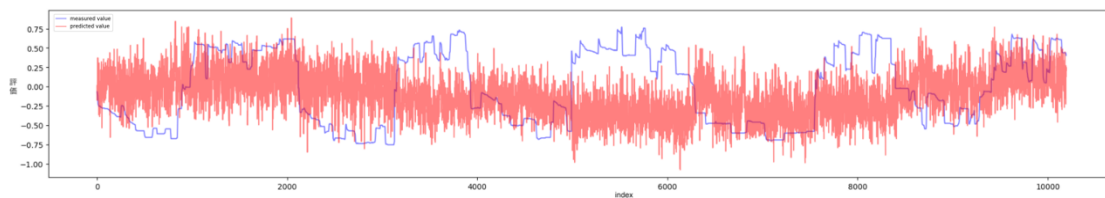


図 A.1.1 外挿について Fold 1 のテストデータに対する推定結果(Arousal)

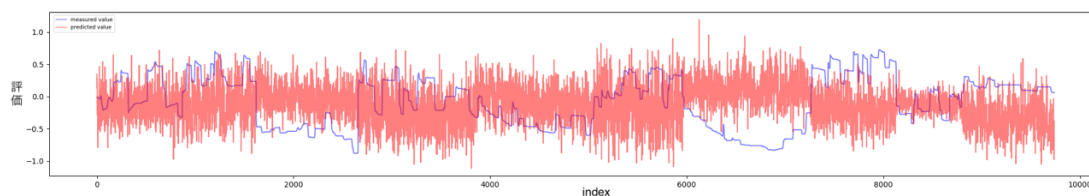


図 A.1.2 外挿について Fold 2 のテストデータに対する推定結果(Arousal)

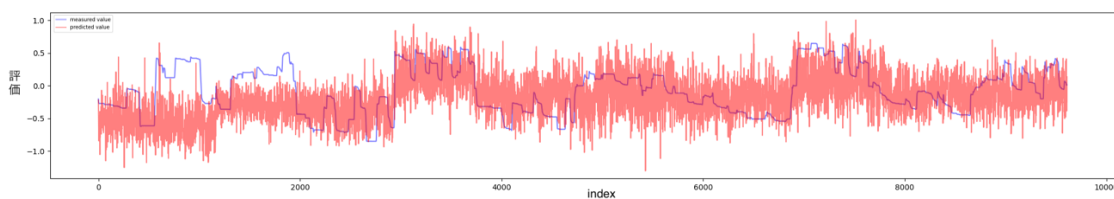


図 A.1.3 外挿について Fold 3 のテストデータに対する推定結果(Arousal)

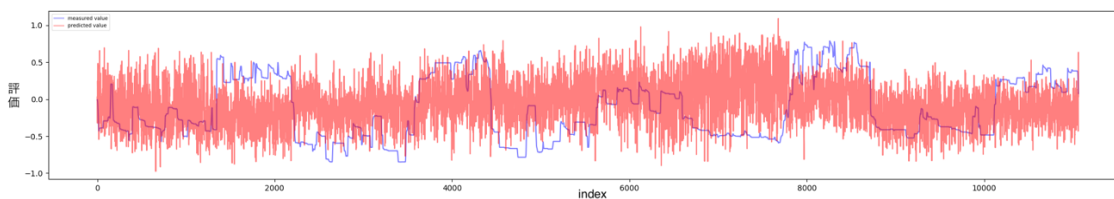


図 A.1.4 外挿について Fold 5 のテストデータに対する推定結果(Arousal)

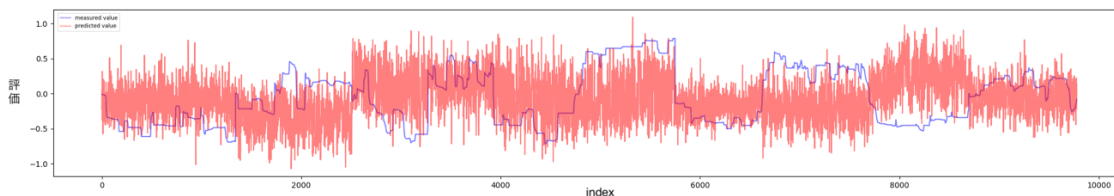


図 A.1.5 外挿について Fold 6 のテストデータに対する推定結果(Arousal)

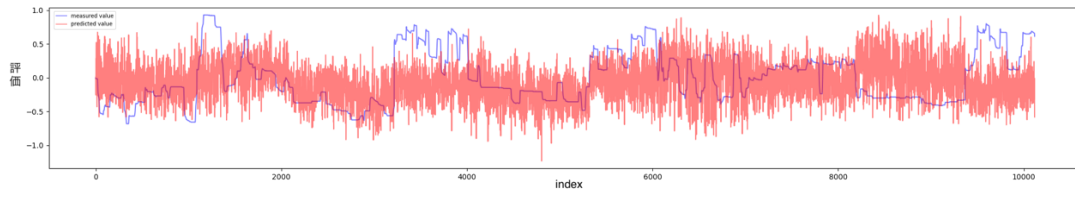


図 A.1.6 外挿について Fold 7 のテストデータに対する推定結果(Arousal)

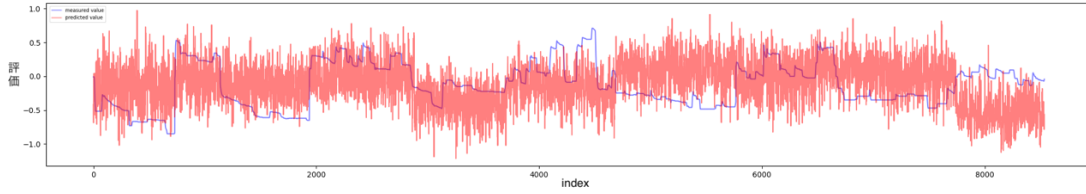


図 A.1.7 外挿について Fold 8 のテストデータに対する推定結果(Arousal)

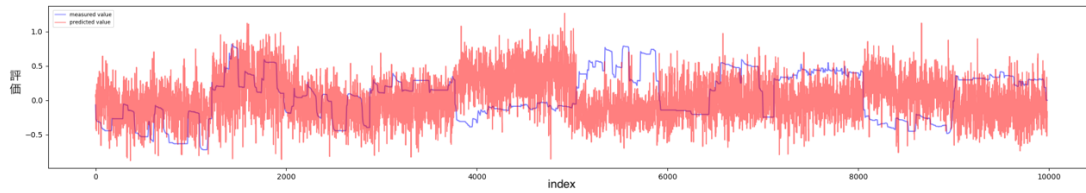


図 A.1.8 外挿について Fold 9 のテストデータに対する推定結果(Arousal)

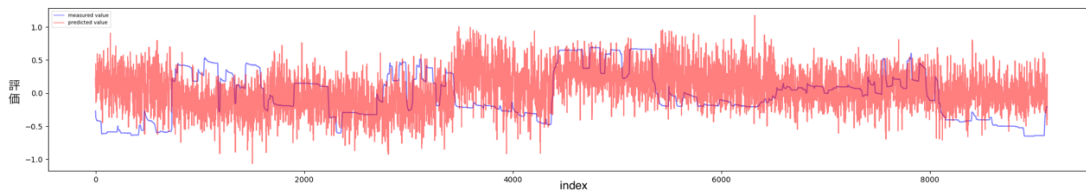


図 A.1.9 外挿について Fold 10 のテストデータに対する推定結果(Arousal)

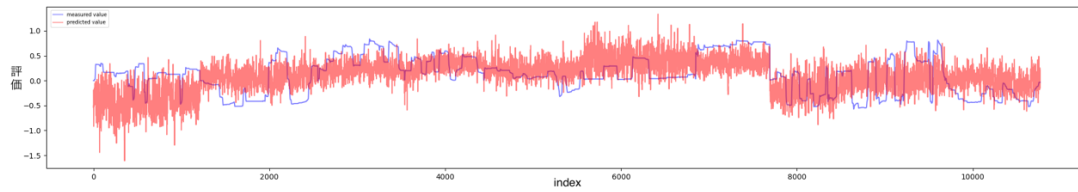


図 A.2.1 外挿について Fold 2 のテストデータに対する推定結果(Valence)

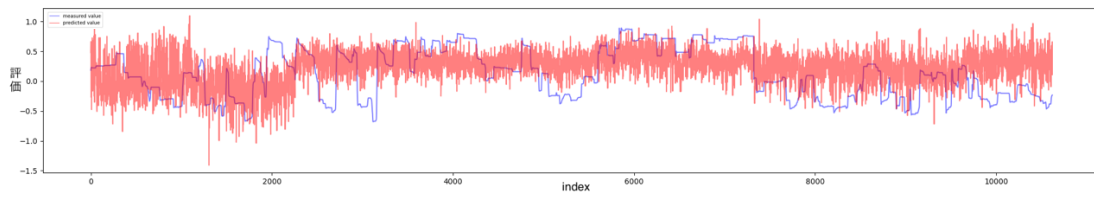


図 A.2.2 外挿について Fold 3 のテストデータに対する推定結果(Valence)

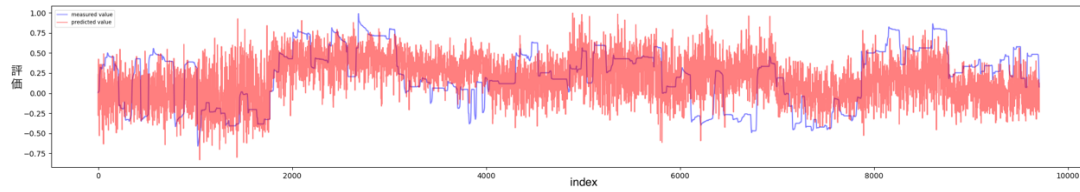


図 A.2.3 外挿について Fold 4 のテストデータに対する推定結果(Valence)

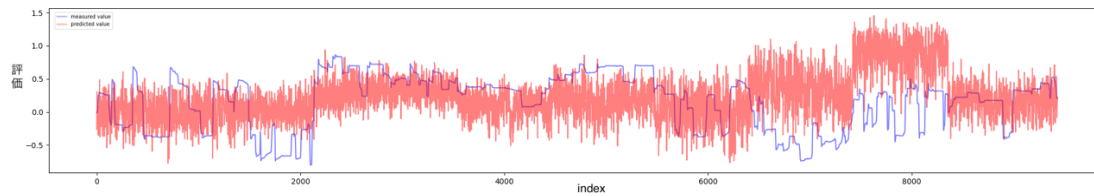


図 A.2.4 外挿について Fold 5 のテストデータに対する推定結果(Valence)

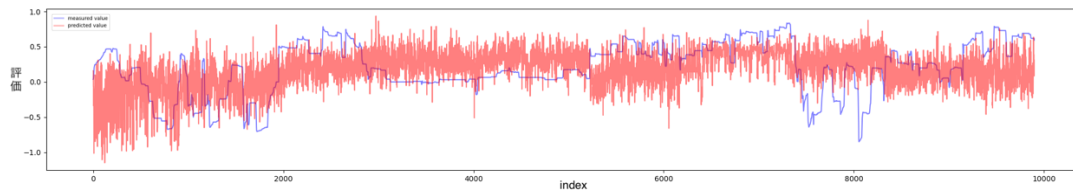


図 A.2.5 外挿について Fold 6 のテストデータに対する推定結果(Valence)

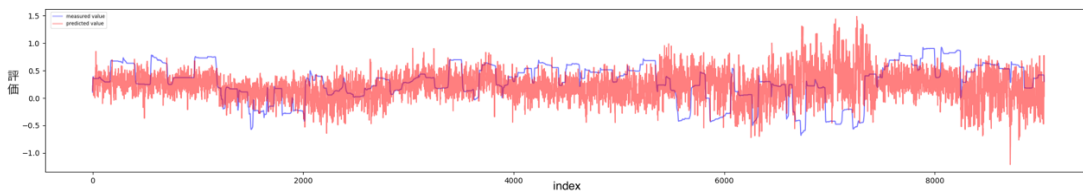


図 A.2.6 外挿について Fold 7 のテストデータに対する推定結果(Valence)

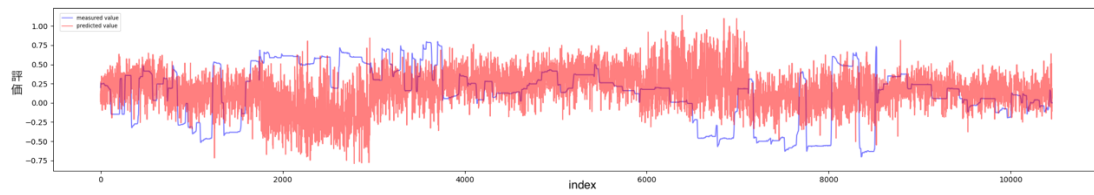


図 A.2.7 外挿について Fold 8 のテストデータに対する推定結果(Valence)

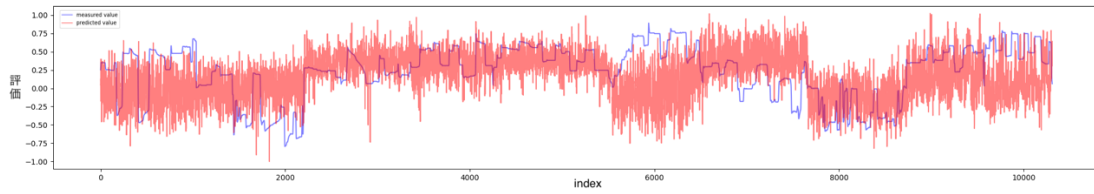


図 A.2.8 外挿について Fold 9 のテストデータに対する推定結果(Valence)

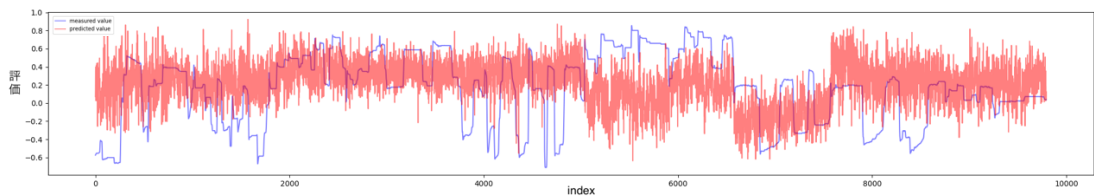


図 A.2.9 外挿について Fold 10 のテストデータに対する推定結果(Valence)

付録 B AutoEncoder で解析した際の model loss

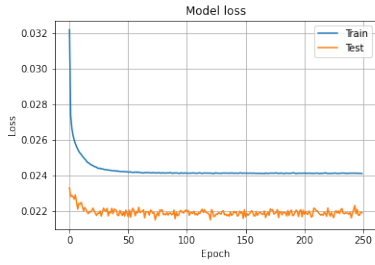


図 B.1.1 4 分類の model loss(Fold 1)

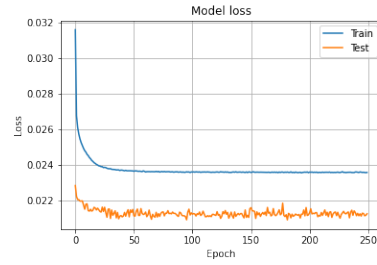


図 B.1.2 4 分類の model loss(Fold 2)

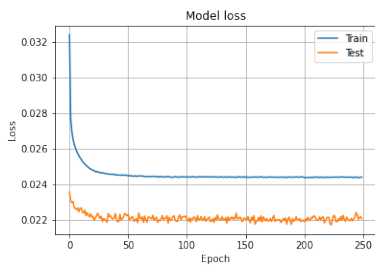


図 B.1.3 4 分類の model loss(Fold 3)

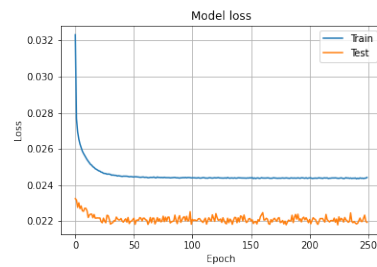


図 B.1.4 4 分類の model loss(Fold 4)

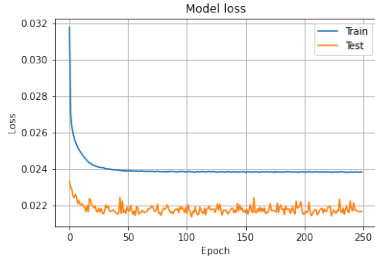


図 B.1.5 4 分類の model loss(Fold 5)

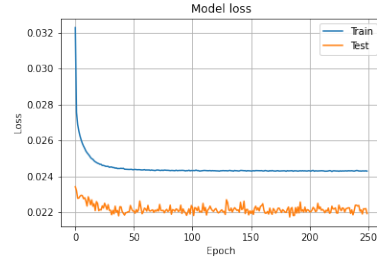


図 B.1.6 4 分類の model loss(Fold 6)

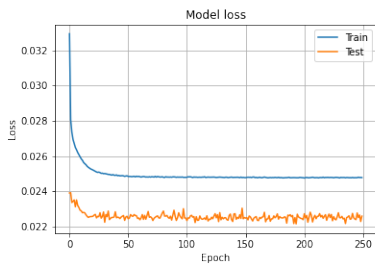


図 B.1.7 4 分類の model loss(Fold 7)

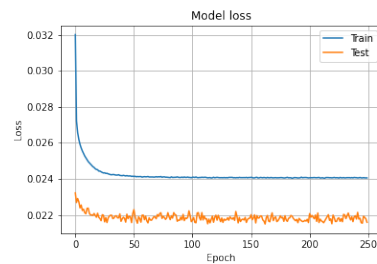


図 B.1.8 4 分類の model loss(Fold 8)

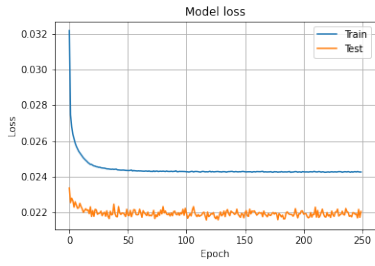


図 B.1.9 4 分類の model loss(Fold 9)

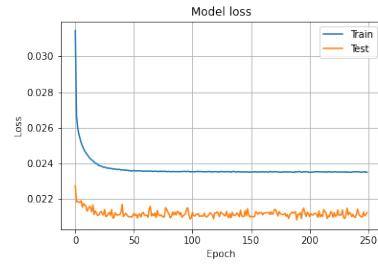


図 B.1.10 4 分類の model loss(Fold 10)

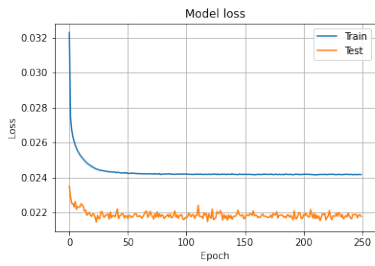


図 B.2.1 「楽しい」の model loss(Fold 1)

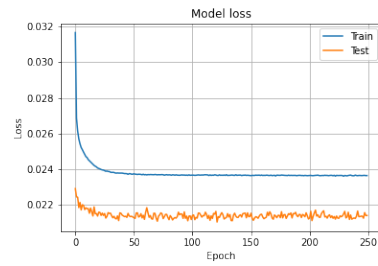


図 B.2.2 「楽しい」の model loss(Fold 2)

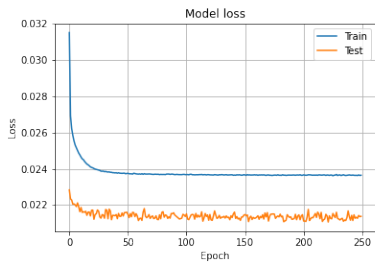


図 B.2.3 「楽しい」の model loss(Fold 3)

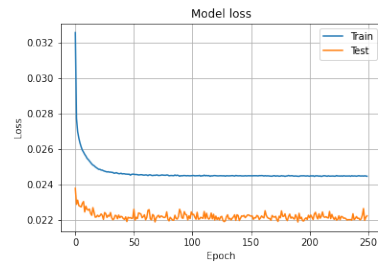


図 B.2.4 「楽しい」の model loss(Fold 4)

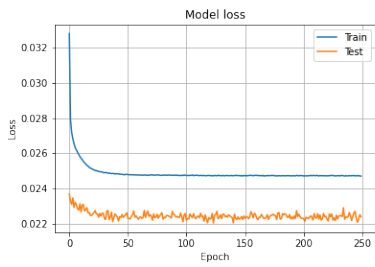


図 B.2.5 「楽しい」の model loss(Fold 5)

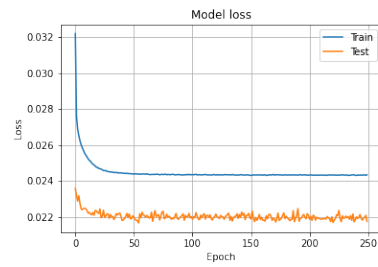


図 B.2.6 「楽しい」の model loss(Fold 6)

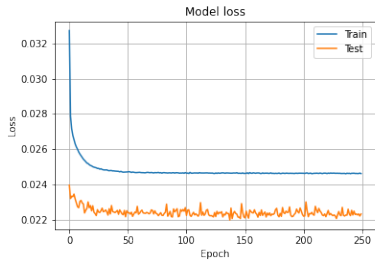


図 B.2.7 「楽しい」の model loss(Fold 7)

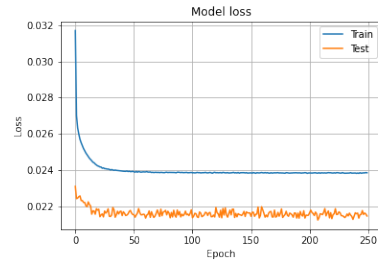


図 B.2.8 「楽しい」の model loss(Fold 8)

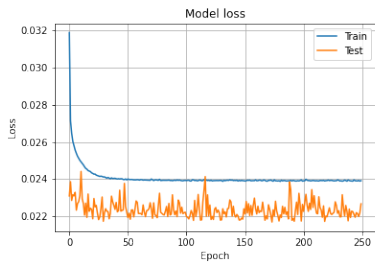


図 B.2.9 「楽しい」の model loss(Fold 9)

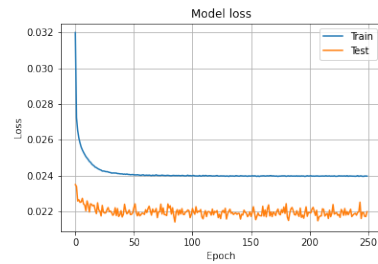


図 B.2.10 「楽しい」の model loss(Fold 10)

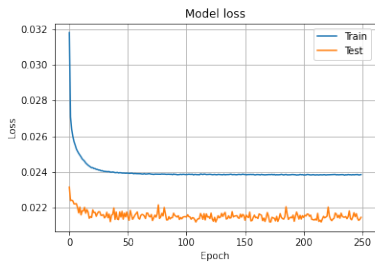


図 B.3.1 「落ち着く」の model loss(Fold 1)

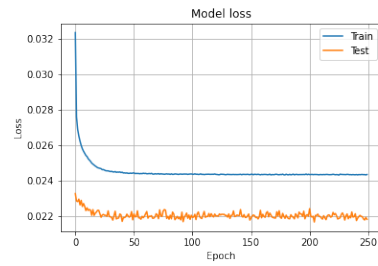


図 B.3.2 「落ち着く」の model loss(Fold 2)

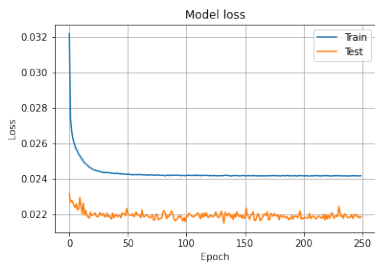


図 B.3.3 「落ち着く」の model loss(Fold 3)

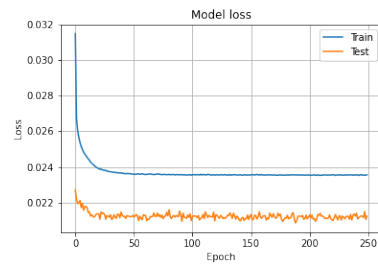


図 B.3.4 「落ち着く」の model loss(Fold 4)

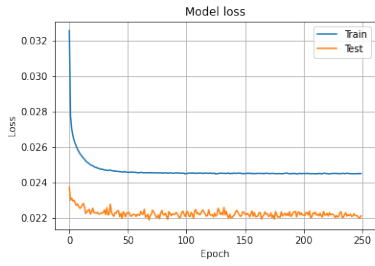


図 B.3.5 「落ち着く」の model loss(Fold 5)

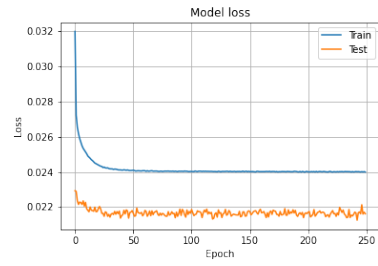


図 B.3.6 「落ち着く」の model loss(Fold 6)

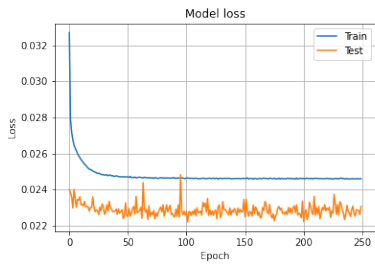


図 B.3.7 「落ち着く」の model loss(Fold 7)

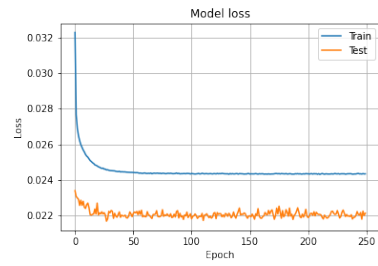


図 B.3.8 「落ち着く」の model loss(Fold 8)

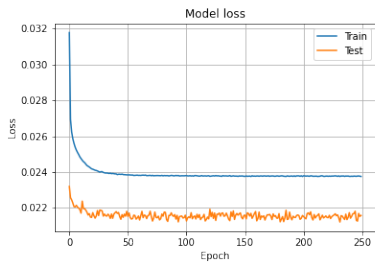


図 B.3.9 「落ち着く」の model loss(Fold 9)

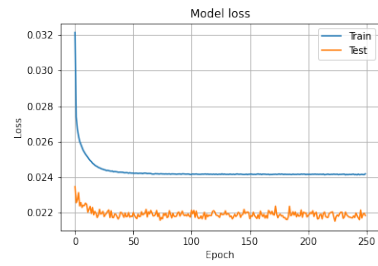


図 B.3.10 「落ち着く」の model loss(Fold 10)

参考文献

- [1] 澁田留奈, 田畑有紀子, 菅沼睦, 亀山渉, “EEG による楽曲聴取者の嗜好の推定に関する研究”, 電子情報通信学会, 2019 年総合大会, H-2-21, 2019 年 3 月
- [2] 田畑有紀子, 澁田留奈, 菅沼睦, 亀山渉, “生体情報を用いた楽曲聴取者の嗜好および楽曲印象に関する研究”, 電子情報通信学会, 2019 年総合大会, H-2-22, 2019 年 3 月
- [3] [3] N. Tanizawa, M. Suganuma and W. Kameyama, “Estimating Music Listener’s Emotion from Bio-signals by using CNN,” IEICE Communication Express, vol.9, no.12, pp.605-609, 2020
- [4] “Spotify: 音楽発見サービス”, <https://www.spotify.com/jp/>, (最終アクセス 2022 年 1 月 18 日)
- [5] 飛松省三, “脳波判読のポイント”, 2006 年 4 月 5 日, <https://www.med.kyushu-u.ac.jp/neurophy/point.pdf>, (最終アクセス 2022 年 1 月 18 日)
- [6] 石田崇, 梅澤克之, 齋藤友彦, 中澤真, 平澤茂一, “プログラミング学習時における簡易脳波計による脳波計測とその分析”, <http://www.waseda.jp/prjngel/material/pub/16/IUSNH16.pdf>, (最終アクセス 2022 年 1 月 18 日)
- [7] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen, "EEG-Based Emotion Recognition in Music Listening", IEEE Transactions on Biomedical Engineering, vol.57, no.7, pp.1798-1806, Jul. 2010
- [8] 笠巻祐二, “I-④ 心電図の波形と正常心電図”, トーアエイヨー, 2013 年, <https://med.toaeiyo.co.jp/contents/ecg/pdf/ecg1-4.pdf>, (最終アクセス 2022 年 1 月 18 日)
- [9] 中川千鶴, “特集③人間工学のための計測手法第 4 部:生体電気現象その他の計測と解析 (5) 1-自律神経系指標の計測と解析-, 人間工学, vol.52, no.1, pp.6-12, 2016, https://www.jstage.jst.go.jp/article/jje/52/1/52_6/_pdf, (最終アクセス 2022 年 1 月 18 日)
- [10] 古川茂人, 米家惇, H.Liao, 柏野牧夫, “眼から読み取る心の動き -Heart-Touching-AI のキー技術”, NTT 技術ジャーナル, vol.28, no.2, pp22-25, Feb. 2016, <https://journal.ntt.co.jp/backnumber2/1602/files/jn201602022.pdf>, (最終アクセス 2022 年 1 月 18 日)
- [11] 鶴飼一彦, “瞳孔・調節・眼球運動計測と映像による生体影響評価”, 日本生体医工学会, BME, Vol.18, No.1, pp.2-7, 2004 年 1 月, https://www.jstage.jst.go.jp/article/jsmbe1987/18/1/18_1_2/_pdf, (最終アクセス 2022 年 1 月 18 日)
- [12] 柏野牧夫, 米家惇, H.Liao, 古川茂人, “身体から潜在的な心を解読するマインドリーダー

- ィング技術”, NTT 技術ジャーナル, vol.26, no.9, pp32-36, Sep. 2014,
<https://journal.ntt.co.jp/backnumber2/1409/files/jn201409032.pdf>, (最終アクセス
2022 年 1 月 18 日)
- [13] WIKIMEDIA COMMONS,
[https://commons.wikimedia.org/wiki/File:International_10-20_system_for_EEG-
MCN.svg](https://commons.wikimedia.org/wiki/File:International_10-20_system_for_EEG-MCN.svg), (最終アクセス 2022 年 1 月 18 日)
- [14] J.M. Girard and A.G.C. Wright. “DARMA: Software for dual axis rating and media
annotation,” Behavior Research Methods, vol.50, no.3, pp.902–909, 2018
- [15] A. Delorme and S. Makeig, “EEGLAB: An open source toolbox for analysis of
single-trial EEG dynamics including independent component analysis,” Journal of
Neuroscience Methods, vol.134, no.1, pp.9-21, 2004.
- [16] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld and M. Tangermann,
“Robust artifactual independent component classification for BCI practitioners,”
Journal of Neural Engineering, vol.11, no.3, 035013, 2014
- [17] “ランダムフォレスト”, 調査・統計用語集, 株式会社日経リサーチ,
<https://www.nikkei-r.co.jp/glossary/id=1599>, (最終アクセス 2022 年 1 月 18 日)
- [18] “scikit-learn”, <https://scikit-learn.org/stable/>, (最終アクセス 2022 年 1 月 18 日)
- [19] “ニューラルネットワークとは?人工知能の基本を初心者向けに解説!”, Udemy メディ
ア, 2017 年 10 月 6 日, <https://udemy.benesse.co.jp/ai/neural-network.html>, (最終ア
クセス 2022 年 1 月 18 日)
- [20] “ディープニューラルネットワーク(DNN: Deep Neural Network)とは?”, atmarkIT,
2019 年 2 月 24 日, <https://www.atmarkit.co.jp/ait/articles/1901/06/news034.html>, (最
終アクセス 2022 年 1 月 18 日)
- [21] “Keras Documentation”, <https://keras.io/>, (最終アクセス 2022 年 1 月 18 日)
- [22] “TensorFlow”, <https://www.tensorflow.org/>, (最終アクセス 2022 年 1 月 18 日)
- [23] “定番の Convolutional Neural Network をゼロから理解する”, DeepAge, 2016 年 11
月 7 日,
https://deepage.net/deep_learning/2016/11/07/convolutional_neural_network.html,
(最終アクセス 2022 年 1 月 18 日)
- [24] “WebAPI | Spotify for Developers”,
<https://developer.spotify.com/documentation/web-api/>, (最終アクセス 2022 年 1 月 18
日)
- [25] “AutoEncoder を可視化しながら試してみる”, AI 研究所, 2017 年 11 月 1 日,
<https://ai-kenkyujo.com/autoencoder/>, (最終アクセス 2022 年 1 月 18 日)
- [26] “オートエンコーダー_Auto Encoder (Vol.21)”, 株式会社システムインテグレータ,
2018 年 6 月 28 日, <https://products.sint.co.jp/aisia/blog/vol1-21>, (最終アクセス 2022

年 1 月 18 日)

[27] “勾配ブースティング回帰木を使用した回帰分析”, 日本電気株式会社, 2021 年 6 月 29 日, <https://jpn.nec.com/hpc/sxauroratsubasa/column/012.html>, (最終アクセス 2022

年 1 月 18 日)

[28] “LightGBM”, <https://lightgbm.readthedocs.io/en/latest/#>, (最終アクセス 2022 年 1 月 18 日)

[29] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta Masanori Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework”, In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2623-2631, 2019, <https://optuna.readthedocs.io/en/stable/index.html>, (最終アクセス 2022 年 1 月 18 日)

謝辞

本研究を進めるにあたり、ご指導及び多くの助言をくださった亀山渉教授、菅沼睦氏に心より感謝し、深くお礼申し上げます。また、共同で実験に取り組んだ榎谷亜美さん始め、本研究に関する助言をくださった亀山研究室の皆様に、心より感謝申し上げます。

研究業績

題目	発表年月	発表	連名者
Estimating Music Listener's Emotion from Bio-signals by using CNN	2020年12月	IEICE Communications Express, 2020, vol.9, no.12, pp.605-609.	菅沼睦, 亀山渉
Estimating Music Listener's Emotion from Bio-signals with EEG Denoising and for Unlearned Music Pieces	2021年10月	2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), pp.75-76, Kyoto, Japan, Oct. 12-15, 2021.	菅沼睦, 亀山渉
生体情報を用いた楽曲聴取者の感情推定における未学習楽曲の推定精度向上に関する検討	2022年3月 (予定)	電子情報通信学会 IMQ/CQ/IE/MVE研究会	菅沼睦, 亀山渉