

早稲田大学院基幹理工研究科情報理工・情報通信専攻

Master's Thesis  
修士論文

Title  
論文題目

楽曲聴取中の基礎律動と楽曲の印象及び聴取楽曲との  
対応に関する研究

A Study on Correspondence between Background Activity while  
Listening to Music and Music or Listener's Impression on Music

学籍番号 Student ID	5116F091-4
氏名 Name	吉田 真嵩 Masataka Yoshida

指導教員 Supervisor	亀山 涉 印 Wataru Kameyama
--------------------	---------------------------

2018年1月30日

2017 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

楽曲聴取中の基礎律動と楽曲の印象及び聴取楽曲との  
対応に関する研究

2018.1.30

吉田 真嵩  
(5116F091-4)

指導教員  
亀山 渉教授

第1章 序論 .....	1
1.1 研究背景・目的 .....	1
1.2 本論文の構成 .....	1
第2章 先行研究 .....	2
2.1 基礎律動 .....	2
2.2 基礎律動を用いた楽曲聴取中の感情推定 .....	2
2.3 脳波を用いた楽曲のジャンル判定 .....	2
2.4 基礎律動を用いたユーザの楽曲に対する楽曲推定 .....	2
2.5 楽曲聴取の順序効果 .....	3
2.6 基礎律動を用いた印象の類似性による楽曲分類 .....	3
第3章 実験およびデータ処理 .....	4
3.1 実験の概要 .....	4
3.2 被験者 .....	4
3.3 使用機材 .....	4
3.4 実験に使用した楽曲 .....	5
3.5 実験手順 .....	10
3.6 アンケート内容 .....	11
3.7 データ処理 .....	11
第4章 アンケートと基礎律動の対応に関する解析 .....	13
4.1 解析の前処理 .....	13
4.2 SVMを用いた検証 .....	13
4.3 ディープラーニングを用いた検証 .....	16
4.4 考察とまとめ .....	25
第5章 聴取楽曲及びジャンルと基礎律動との対応に関する解析 .....	26
5.1 オートエンコーダ .....	26
5.2 オートエンコーダを用いた学習 .....	26
5.3 k-means 法によるクラスタリング .....	53
5.4 楽曲ごとの比較 .....	54
5.5 ジャンルごとの比較 .....	61
5.6 同一音源を複数回聴取した場合の比較 .....	68
5.7 クラスタリング結果の定量的な評価手法の検討 .....	74
5.8 まとめと考察 .....	79

第 6 章 結論 .....	80
6.1 まとめ .....	80
6.2 今後の課題 .....	81
図一覧 .....	82
表一覧 .....	86
謝辞 .....	87
参考文献 .....	88
付録 .....	89
研究業績 .....	90



# 第1章 序論<sup>1</sup>

## 1.1 研究背景・目的

今日では、Apple Music[1]や、Line Music[2]などの様々な音楽配信サービスが普及した。これによって、ユーザが手軽に入手できる楽曲が増え、インターネット上で音楽に触れる機会が大きく増加した。このことによって、たくさんの音楽を聴取できる反面、ユーザが自分の嗜好に合った楽曲を探すことが難しくなっているという問題点があげられる。

この問題を解決するために、ユーザの購入、聴取履歴などを利用して、コンテンツの中から個人の嗜好などにそったコンテンツを推薦する、協調フィルタリングなどのシステムが存在する。しかし、現在のコンテンツ推薦システムはユーザの履歴を元に推薦するシステムであるため、楽曲を聴取した時の感じ方や印象などの個人間での差異が考慮されにくく推薦精度が低い場合がある。そこで現在、生体情報を用いたコンテンツ推薦システムの研究が注目されている。

本研究は、生体情報を利用してユーザ個人の情報を用いることで、より良いコンテンツ推薦システムを構築することを目標としている。その中でも、本論文では音楽聴取中の基礎律動をもちいて、楽曲聴取者が楽曲に対して持った印象を推定することを目標とした。

## 1.2 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および目的を示す。第2章では音楽と基礎律動に関する先行研究を示す。第3章では、実験の概要、被験者、使用機材、使用した音楽コンテンツ、アンケート内容、アンケート結果および取得データの処理を示す。第4章では、取得データとアンケート結果の対応に関して行った解析とその結果及び考察について示す。第5章では、取得データと聴取楽曲及びジャンルとの対応に関して行った解析とその結果及び考察について示す。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。

---

<sup>1</sup> この章は自著文献[16]の内容を含む

## 第2章 先行研究

### 2.1 基礎律動

脳波とは、脳の活動によって生じる電気活動を、電極を用いて記録したものである。計測する部位によって検出される脳波パターンは異なるため、国際 10-20 法で計測部位を国際的に統一している。脳波のうち、持続的に表れる特定の周波数帯域を基礎律動という。基礎律動は周波数帯域ごとに分類され、 $\alpha$ 波 8-13[Hz]、 $\beta$ 波 14-24[Hz]、 $\theta$ 波 4-7[Hz]、 $\delta$ 波 1-3[Hz]などがあり、一般的に $\alpha$ 波は安静時に現れやすく、 $\beta$ 波は緊張時に現れやすいといわれている[3]。

### 2.2 基礎律動を用いた楽曲聴取中の感情推定

Lin らは、楽曲聴取時の左右の複数の電極の脳波データに対してフーリエ変換を行って得たパワースペクトルの差のデータセットを用いて、楽曲聴取中のユーザの感情推定が可能であると示唆している[4]。

### 2.3 脳波を用いた楽曲のジャンル判定

楽曲聴取時の脳波に主成分分析およびカーネル主成分分析を行い、脳波の特徴を抽出することで、聴取した楽曲のジャンルによって異なるパターンが現れることが小川らによって報告されている[5]。このことから、脳波の特徴からユーザが聴取している楽曲ジャンルの推定が可能であると示唆される。しかし、同ジャンルの楽曲でも脳波のパターンが異なる場合がある。これは、同じジャンルの楽曲でも、ユーザの印象等により異なる場合があること、また、脳波のパターンにも個人差があることなどが要因となっていると考えられる。

### 2.4 基礎律動を用いたユーザの楽曲に対する楽曲推定

筆者らは、楽曲聴取中の基礎律動データに対してフーリエ変換を行って得られたパワースペクトルを用いて、SVM による分類を行うことで、楽曲聴取者の楽曲に対する印象推定の可能性を報告した[6]。内挿における分類予測は比較的高い精度で行えるのに対し、外挿による予測がチャンスレベル程度の精度に落ちてしまうことが課題となって

いる。

## 2.5 楽曲聴取の順序効果

楽曲聴取中の基礎律動データを用いた、楽曲の聴取順序の推定の可能性が報告されている。澁田らは SVM を用いて、基礎律動データを楽曲の聴取順序ごとに分類する手法を提案している。提案された手法を用いることで、内挿では高い精度で分類でき、外挿での予測精度はチャンスレベルを超えた精度が得られたと報告している[7]。

## 2.6 基礎律動を用いた印象の類似性による楽曲分類

楽曲聴取中の基礎律動データを用いて、印象の類似性を元に楽曲の分類ができる可能性があることが報告されている。三木らは SVM を用いて、基礎律動データを楽曲の印象の類似性ごとに分類する方法を提案している。提案された手法を用いることで内挿では高い予測精度で分類できるが、外挿の予測精度が低いことが課題であると報告されている[8]。

## 第3章 実験およびデータ処理

### 3.1 実験の概要

本研究の実験では、歌詞に日本語が使用されている様々な楽曲を被験者に聴取させ、楽曲聴取中の基礎律動を取得する実験を行った。また、被験者にアンケート調査を実施し、楽曲に対する印象も同時に調べた。

先行研究[6]では、内挿予測ではチャンスレベルを超えて識別可能であると報告されている。しかし、外挿による分類ができていないことが課題となっている。実際に音楽鑑賞時に利用することを考えた際に、外挿の精度を改善することは不可欠である。

一般的に機械学習によるデータの分類を行う際は、データ量を増やすことで様々なパターンのデータを学習することができ、精度が改善することが期待される。そこで、データ不足を解消するため、実験に使用する楽曲数を先行研究[6]よりも増加させ、200楽曲にして実験を行った。

### 3.2 被験者

被験者は日本国籍をもつ男性1名であり、年齢は24歳だった。

被験者は実験時に覚醒状態にあった。

### 3.3 使用機材

基礎律動の取得には、Interaxon社のMuseを用いた。Museは簡易脳波計であり、基準電極はFPzである。

今回の実験で使用したプリセットおよび測定点の情報を次に示す[9]。また、電極の位置を示した図を図3.1に示す。計測した電極の位置を青で、基準電位の位置を赤で示した。

【プリセット情報】

電極位置：TP9、AF7、AF8、TP10

EEGデータ：500Hz、16bit

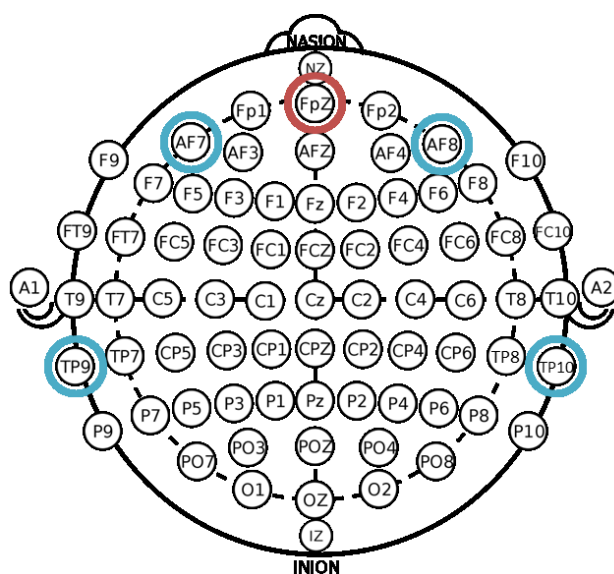


図 3.1 muse の電極位置

([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) より引用し編集を行った、元画像はパブリックドメインである。)

### 3.4 実験に使用した楽曲

本実験では、歌詞に日本語が使用されている楽曲 200 曲を使用して実験を行った。使用した楽曲の一覧を表 3.1 から表 3.4 に示す。実験時には、それぞれの楽曲のもっとも特徴的であると考えられる部分 90 秒間を抜き出して実験に使用した。以降、本稿ではこれを音源と呼ぶ。

表 3.1 実験に使用した楽曲 1

音源番号	アーティスト名	楽曲名
1	Cyntia	閃光ストリングス
2	Galneryus	Future Never Dies
3	陰陽座	野衾忍法帖
4	Mary's Blood	Marionette
5	Anthem	Black Empire
6	supercell	perfect day
7	絢香	手をつなごう
8	Chris Hart	僕が一番欲しかったもの
9	玉置浩二	Lion(Album Ver.)
10	カノン	フタリ
11	細川たかし	北緯50度
12	坂本冬美	また君に恋してる
13	舟木一夫	天城越え
14	八代亜紀	おんな港町 2002バージョン
15	福田こうへい	門出
16	クリープハイブ	リバーシブルー
17	HY	あなたを想う風
18	Lightning	Raise the sun
19	Linked Horizon	紅蓮の弓矢
20	チャットモンチー	風吹けば恋
21	ゆず	いつか
22	スキマスイッチ	全力少年
23	Aldious	Dominator
24	新沼謙治	風列車
25	奥華子	ガーネット
26	[Alexandros]	ワタリドリ
27	Babymetal	Road of Resistance
28	NICO Touches the Walls	B.C.G
29	一青窈	もらい泣き
30	茅原実里	境界の彼方
31	フジファブリック	若者のすべて
32	喜多村英梨	incomplete
33	Loudness	Crazy Doctor
34	ゲスの極み乙女	キラールール
35	大川栄策	雨の港
36	Unison Square Garden	KID, I LIKE QUARTED
37	Gackt	君が追いかけた夢
38	東京事変	金魚の箱
39	tetsuya	REVERSE
40	Triceratops	Fever
41	B'z	Seventh Heaven
42	Ken	solitary stroll
43	シド	Smile
44	宇多田ヒカル	DISTANCE
45	L'Arc~en~Ciel	Link
46	きゃりーぱみゅぱみゅ	にんじゃりばんばん
47	森山直太朗	生きてることが辛いなら
48	天野月子	鳥
49	L'Arc~en~Ciel	暴食・lose control
50	Acidman	造花が笑う

表 3.2 実験に使用した楽曲 2

音源番号	アーティスト名	楽曲名
51	VAMPS	Memories
52	竹澤汀	東京タワー
53	FUZZY CONTROL	罨
54	Judy and Mary	そばかす
55	はいだしょうこ	おもちゃのチャチャチャ
56	miwa	friends ~君が笑えば~
57	椎名林檎	眩暈
58	The Back Horn	ジョーカー
59	amazarashi	千年幸福論
60	Crystal Kay	何度でも
61	Greeeen	キセキ
62	高橋優	WC
63	La'cryma Christi	未来航路
64	HYDE	Angel's Tale
65	Dreams Come True	うれしい! たのしい! 大好き! (Everlasting Version)
66	鬼束ちひろ	This Silence Is Mine
67	Sex Machinguns	Tekken II
68	Yoshii Lovinson	Call Me
69	美空ひばり	愛燦燦
70	HY	366日
71	Gackt	月の詩
72	中島みゆき	地上の星
73	Fence of Defence	Sara
74	Babymetal	Karate
75	森山直太郎	若者たち
76	May J	Let it Go
77	はいだしょうこ	故郷
78	ポルノグラフィティ	音のない森
79	Bump of Chicken	乗車権
80	Sons of All Pussys	Paradise
81	L'Arc~en~Ciel	Dive To Blue
82	Andymori	City Lights
83	斉藤和義	雨宿り
84	Home Made 家族	アイコトバ
85	syrup 16g	負け犬
86	indigo La End	緑の少女
87	GODIEGO	銀河鉄道999
88	nobodyknows+	ココロオドル
89	鬼束ちひろ	流星群
90	Androp	Yeah!Yeah!Yeah!
91	RHYMSTER	人間交差点
92	氣志團	結婚闘魂行進曲マブダチ
93	スーパーカー	New Young City
94	ももいろクローバーZ	走れ!
95	Amazarashi	ムカデ
96	七尾旅人	サーカスナイト
97	八代亜紀	舟歌
98	Seamo	ルパン・ザ・ファイアー
99	クラムボン	yet
100	韻シスト	Bird

表 3.3 実験に使用した楽曲 3

音源番号	アーティスト名	楽曲名
101	浦島太郎 (桐谷健太)	海の声
102	Syrup 16g	She was beautiful
103	RIP SLYME	熱帯夜
104	工藤静香	嵐の素顔
105	BEGIN	島唄
106	氣志團	One night carnival (Album version)
107	原田郁子	ユニコーン
108	中西保志	最後の雨
109	LUNA SEA	ROSIER
110	岡村靖幸 & 小出祐介	愛はおしゃれじゃない
111	ガガガSP	国道二号線
112	玉置浩二	闇をロマンスにして
113	Perfume	Dream Fighter
114	フジファブリック	Strawberry Shortcakes
115	Galneryus	Destiny
116	CHARCOAL FILTER	サヨナラ
117	ムック	リブラ
118	凜として時雨	ターボチャージャーON
119	Rebecca	フレンズ
120	スーパーカー	Low-Down (Live Scene)
121	絢香	Real Voice
122	本田美奈子	見上げてごらん夜の星を
123	L'Arc~en~Ciel	Drink it Down
124	KOKIA	You
125	Acidman	star train
126	鬼束ちひろ	螺旋
127	Koji Nakamura	Shutoko
128	SUEMITSU & THE SUEMITH	Astaire
129	藍坊主	水に似た感情
130	摩天楼オペラ	輝きは閃光のように
131	布袋寅泰	RUSSIAN ROULETTE
132	hide	ROCKET DIVE
133	Spicy Chocolate	ずっと feat.HAN-KUN & TEE
134	オフコース	時に愛は
135	絢香	三日月
136	頑張れ! Victory	夢のつづき
137	Segare	枯木に花
138	新沼謙治	俺の昭和が遠くなる(シングルバージョン)
139	しほん	見えてない私
140	ゲスの極み乙女。	ゲスな三角関係
141	orange pekoe	やわらかな夜
142	RADWIMPS	いいんですか?
143	八神純子	1年と10秒の交換
144	リュウグウノツカイ	琉球の空
145	夜の本気ダンス	Bltch
146	AI	みんなが英雄 (フルバージョン)
147	河口京吾	桜(2013 NEW REMASTER)
148	玉置浩二	花咲く土手で
149	大石まどか	命、燃えて
150	TEE	ベイビーアイラブユー



表 3.4 実験に使用した楽曲 4

音源番号	アーティスト名	楽曲名
151	猿岩石	白い雲のように
152	羊毛とおはな	ずっとずっとずっと
153	クレイジーケンバンド	家に帰ろうよ
154	小田純平	250パーセントの愛でつつんで
155	Base Ball Bear	ファンファーレが聞こえる (Album Mix)
156	indigo La End	sweet spider
157	片平里菜	始まりに
158	鬼束ちひろ	私とワルツを
159	シャ乱Q	シングルベッド
160	THE YELLOW MONKEYS	楽園
161	BONNIE PINK	A Perfect Sky
162	工藤静香	雨夜の月に
163	ORANGE RANGE	花
164	ANTHEM	Sailing
165	松城和生	ひとりの冬に
166	Aqua Time z	虹
167	竹仲絵里	サヨナラ サヨナラ
168	One Ok Rock	Wherever you are
169	LIQUID CHORD	星の彼方
170	amazarashi	アポロジー
171	Ryo from Orange Range	Tim Don!-Don! Feat. DJ KEIN
172	Dir en Grey	ain't afraid to die
173	内田克彦	もつ心
174	東京事変	絶対値対相対値
175	moumoon	トモシビ
176	Barbee Boys	もォやだ!
177	鬼束ちひろ	BORDERLINE
178	Mary's Blood	Bite the Bullet
179	Perfume	ナチュラルに恋して
180	カノン	My Destiny
181	WANDS	世界が終わるまでは・・・
182	ムック	茫然自失
183	Che'Nelle	Happiness
184	二代目アニメタル	ライオン
185	山本あき	大阪さみしがり
186	Cross Vein	forget-me-not
187	走裕介	北の傷歌
188	シド	必要悪
189	HOWL BE QUIET	MONSTER WORLD
190	ポルノグラフィティ	渦
191	Creature Creature	パラダイス
192	中島美嘉	雪の華
193	ZARD	運命のルーレット廻して
194	凛として時雨	Hysteric phase show
195	カラスは真っ白	ヒズムリアリズム
196	それでも世界が続くなら	弱者の行進
197	moumoon	Sunshine Girl
198	Ken	Relax Over
199	Mr.Children	掌
200	Galneryus	Everlasting

### 3.5 実験手順

実験は次の手順で行った。

- ① 被験者に脳波計を装着する
- ② 音源を1つ聴取する
- ③ アンケートに回答する
- ④ ②、③を繰り返す

1日当たり、5音源から30音源の聴取を行い、全200音源の聴取を行った。

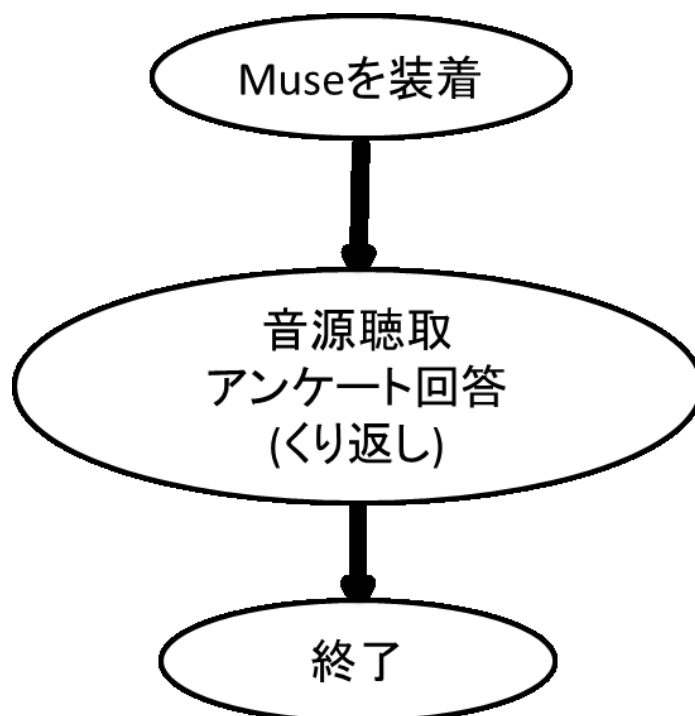


図 3.2 実験の流れ

## 3.6 アンケート内容

アンケートには次の質問項目を設けた。

- ・ 年齢
- ・ 性別
- ・ 普段の音楽鑑賞の頻度(5段階)
- ・ 好みの音楽のジャンル
- ・ 好みの音楽アーティスト
- ・ 実験に用いた各楽曲の認知度合(3段階)
- ・ 実験に用いた各楽曲の嗜好性(5段階)
- ・ 実験に用いた各楽曲の印象評価(SD法)

印象評価の手法としてSD法を用いた。SD法とは、相対する意味の形容詞対を用意し、その間を何段階かに分割し測定する方法である。本実験で用いた評価指標を表3.5に示す。実験では表3.5の評価指標を5段階に分けて被験者に評価してもらった。

表 3.5 実験で使用した評価指標(自著文献[16]より引用)

明るい	暗い
重い	軽い
柔らかい	硬い
安定	不安定
澄んだ	にごった
滑らか	歯切れのよい
激しい	穏やか
厚い	薄い

## 3.7 データ処理

### 3.7.1 基礎律動データ

解析の前処理として、脳波計から出力された電圧値に対し、高速フーリエ変換を施した。フーリエ変換は、1秒の窓を0.05秒ずつスライドさせる形で行い、電極ごとに1-45[Hz]のパワースペクトルを1[Hz]ごとに算出した。なお、被験者が楽曲に集中するまでに時間が掛かる可能性を考慮し、聴取した90秒のうち、中間部の15秒から75秒

までの間の 60 秒間を解析対象とした。高速フーリエ変換を行ったデータのうち 1-45Hz を解析対象とした。FFT 後のデータは以下のように成形した。

$$X = \begin{bmatrix} f_{TP9\ 1} & f_{TP9\ 2} & & \dots \\ f_{AF7\ 1} & \ddots & & \\ f_{AF8\ 1} & & f_{ct} & \dots \\ f_{TP10\ 1} & f_{TP10\ 2} & & \dots \end{bmatrix} \quad (1)$$

$$f_{ct} = \begin{bmatrix} d_{1\ ct} \\ d_{2\ ct} \\ \vdots \\ d_{h\ ct} \\ \vdots \\ d_{45\ ct} \end{bmatrix} \quad (2)$$

$f_{ct}$ の $c$ 、 $t$ に関して、 $c$ は脳波計の電極、 $t$ は時間、つまり、何個目のデータであるかを表す。 $f_{ct}$ には、 $t$ 番目に出力された、電極 $c$ における 1-45[Hz]のパワースペクトルデータが含まれている。また、 $d_{h\ ct}$ は $t$ 番目に出力された、電極 $c$ における $h$ [Hz]のパワー値が含まれている。

### 3.7.2 アンケート集計

回答してもらったアンケートの印象評価の質問対し、表 3.2 の各印象項目の右側を 1、左側を 5 として得点化を行ったものを用意した。得点化したアンケートを、3 点、4 点以上、2 点以下の 3 値に変換したものも用意した。これ以降アンケート結果は 3 値に変換したものを使用する。印象評価指標の一つである「明るいー暗い」を例にとって、どのように変換したかを示す図を図 3.3 に示す。

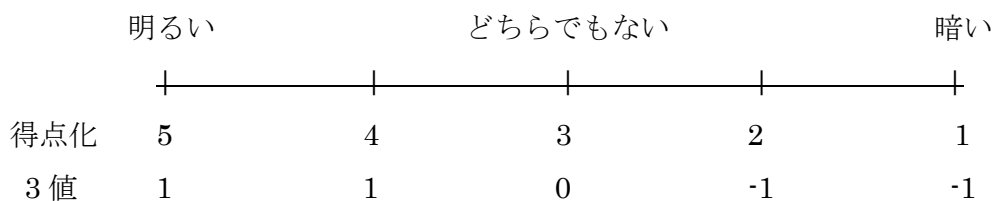


図 3.3 アンケート集計

## 第4章 アンケートと基礎律動の対応に関する解析

この章では、基礎律動データとアンケート結果の対応に関する解析を行った。先行研究[6]では SVM を用いた内挿において、基礎律動データを表 3.5 の各印象評定語の形容詞対を正解データとして、チャンスレベルを超えた精度で分類できると報告している。その一方で、外挿に課題が残っている。外挿予測の課題を解決するために以下の検証を行った。

### 4.1 解析の前処理

本実験では Muse を使用して基礎律動を記録した。Muse は頭部の皮膚上に電極を設置し計測を行う。その際の電極の設置状態により、記録される電位の強弱が異なる可能性がある。これがノイズとなり、分類精度の低下の原因となることが考えられるため、3.7.1 で得られたデータに対して正規化を施した。この章では正規化は、各電極の単位時間当たりのパワー値の総和で各周波数のパワー値を割るという手法で行った。

### 4.2 SVM を用いた検証

SVM を用いて基礎律動データの学習を行い、外挿の予測精度の検証を行った。SVM は検証 1 では R のパッケージである Kernlab[10]を、そのほかでは libSVM[11]を用いた。なお、すべての検証でチャンスレベルは 0.3333 である。

#### 4.2.1 検証 1

アンケートの回答結果通りに基礎律動データが分類できるかどうかの検証を行った。正規化を施した基礎律動データを入力データとし、アンケート結果を正解データとして利用した。

200 音源のデータを音源ごとにランダムに並べ替え、20 音源分取り出しテストデータとし、残りの 180 音源分を学習データとして使用し、分類を行う試行を 10 回行った。

印象評価項目の形容詞対ごとに 10 回の分類精度の平均をまとめたものを表 4.1 に示す。印象評定項目「厚い-薄い」以外はチャンスレベルを超えているものの、最も分類精度がよいもので「安定-不安定」の 0.4683 と低い精度に留まった。

表 4.1 各印象評定語に対する SVM での平均分類精度

印象評定項目	平均分類精度
明るい-暗い	0.3415
重い-軽い	0.4415
柔らかい-硬い	0.4487
安定-不安定	0.4683
澄んだ-濁った	0.4259
滑らか-歯切れのよい	0.4638
激しい-穏やか	0.3714
厚い-薄い	0.3266

## 4.2.2 検証 2

4.2.1 の検証では 1 つの入力データのベクトル当たりの時間方向の範囲が 1 秒分であり、基礎律動の時間方向の変動を考慮できていないという欠点が存在する。そのため、1 つのベクトルデータに時間方向に幅を持たせたデータセットを作成し、分類精度が向上するかどうかの検証を行った。

以下の  $Y$  のように入力データを形成した。 $F_n$  は FFT を施した際に  $n$  番目に得られた 1-45[Hz] のパワースペクトルを 4 チャンネル分結合したものを表す。なお、SVM に入力する際に、 $Y$  に対して Z-Score による正規化を施した。SVM 実行時に正解データとしてアンケート結果の印象評定項目「明るい-暗い」を用いた。

$$F_t = \begin{bmatrix} f_{TP9 t} \\ f_{AF7 t} \\ f_{AF8 t} \\ f_{TP10 t} \end{bmatrix} \quad (3)$$

$$Y = \begin{bmatrix} F_1 & F_2 & \dots \\ F_2 & F_3 & \dots \\ \vdots & \vdots & \\ F_n & F_{n+1} & \dots \end{bmatrix} \quad (4)$$

$n=5$  としてこのデータを用い、4.2.1 の時と同様に、200 音源のデータを音源ごとにランダムに並べ替え、20 音源分取り出しテストデータとし、残りの 180 音源分を学習データとして使用し、分類を行う試行を 10 回行った。

SVM の平均分類精度は 0.3318 であった。4.2.1 の印象評定項目「明るい-暗い」と比較しても、分類精度は大きく変化せず、低い分類精度に留まった。

### 4.2.3 検証 3

4.2.1 および 4.2.2 では、入力データとして 1-45[Hz]の 1[Hz]ごとのパワースペクトルを正規化したものを入力データとして使用していた。一方で、2.1 で示した $\alpha$ 波や、 $\beta$ 波のように、基礎律動を用いる際には特定の周波数帯域を使用する場合も多い。そこで、1-45[Hz]のパワースペクトルをいくつかの周波数帯に分割して検証を行った。

以下の $X'$ のようにデータ処理を行い、1-45[Hz]のパワースペクトルをいくつかの周波数帯に分割した。なお以下の式で、 $w$ は 45 の約数である。

$$f'_{ctw} = \begin{bmatrix} \frac{d_{1ct} + \dots + d_{wct}}{w} \\ \frac{d_{(w+1)ct} + \dots + d_{2wct}}{w} \\ \vdots \\ \frac{d_{(46-w)ct} + \dots + d_{45ct}}{w} \end{bmatrix} \quad (5)$$

$$X' = \begin{bmatrix} f'_{TP9\ 1\ w} & f'_{TP9\ 2\ w} & \dots & \dots \\ f'_{AF7\ 1\ w} & \ddots & & \dots \\ f'_{AF8\ 1\ w} & & f'_{ctw} & \dots \\ f'_{TP10\ 1\ w} & f'_{TP10\ 2\ w} & & \dots \end{bmatrix} \quad (6)$$

なお、SVM に入力する際に、 $X'$ に対して Z-Score による正規化を施した。SVM 実行時に正解データとしてアンケート結果の印象評定項目「明るい-暗い」を用いた。

4.2.1 の時と同様に、200 音源のデータを音源ごとにランダムに並べ替え、20 音源分取り出しテストデータとし、残りの 180 音源分を学習データとして使用し、分類を行う試行を 10 回行った。

$w=3,5,15$ と変化させた場合の SVM の平均分類精度を表 4.2 に示す。

表 4.2  $w$ を変化させたときの SVM の平均分類精度

$w$	平均分類精度
3	0.3388
5	0.3585
15	0.3565

表 4.2 のいずれの場合でもチャンスレベル程度の精度に留まった。基礎律動データをいくつかの周波数帯に分割しても分類精度に大きな影響は見られなかった。

### 4.2.4 検証 4

検証 4 では 4.2.3 で行った 1-45[Hz]のパワースペクトルをいくつかの周波数帯域に

分ける処理を行った後に、4.2.2 で行った 1 つのベクトルデータに時間方向の幅を持たせる処理( $n=5$  で固定)を行い、分類精度に向上が見られるかどうかの検証を行った。具体的には以下の $Y'$ のように入力データを形成した。

$$F'_{tw} = \begin{bmatrix} f'_{TP9tw} \\ f'_{AF7tw} \\ f'_{AF8tw} \\ f'_{TP10tw} \end{bmatrix} \quad (7)$$

$$Y' = \begin{bmatrix} F'_1 & F'_2 & \dots \\ F'_2 & F'_3 & \dots \\ \vdots & \vdots & \\ F'_5 & F'_6 & \dots \end{bmatrix} \quad (8)$$

なお、SVM に入力する際に、 $Y'$ に対して Z-Score による正規化を施した。SVM 実行時に正解データとしてアンケート結果の印象評定項目「明るい-暗い」を用いた。

4.2.1 の時と同様に、200 音源のデータを音源ごとにランダムに並べ替え、20 音源分取り出しテストデータとし、残りの 180 音源分を学習データとして使用し、分類を行う試行を 10 回行った。

$w=3,5,15$ と変化させた場合の SVM の平均分類精度を表 4.2 に示す。

表 4.3  $w$ を変化させたときの SVM の平均分類精度( $n=5$  で固定)

$w$	平均分類精度
3	0.3328
5	0.3499
15	0.3553

$w=3,5,15$ のいずれの場合もほぼチャンスレベルの精度に留まり、本検証で行った、前処理の有効性は確認できなかった。

### 4.3 ディープラーニングを用いた検証

SVM を用いた検証で高い分類精度が得られなかったため、機械学習の方法を変更し、ディープラーニングを用いた検証を行った。ディープラーニングには Keras[12]を使用した。Keras のバックエンドは、Tensorflow[13]を使用した。以下では、正規化を施したパワースペクトルを入力データとし、正解データとしてアンケート結果の印象評定項目「明るい-暗い」を用いた。学習時にはテストデータとして音源番号 1-20 を聴取している時の基礎律動データをテストデータとして使用し、残りの音源を聴取している時の基礎律動データを学習データとして使用した。なお、すべての検証で、チャンスレベル



は 0.3333 である。

### 4.3.1 検証 5

6 層のネットワークを構成し、基礎律動データの分類を試みた。

学習時の層構造を表 4.4 に示す。

表 4.4 検証 5 におけるネットワーク構造

活性化関数	Sigmoid(最後の層の後のみ Softmax)
最適化アルゴリズム	Adam
層数(ノード数)	6(180,900,400,200,50,3)
トレーニングエポック数	500
損失関数	categorical_crossentropy

学習結果(分類精度)を図 4.1 に示す。図中の凡例は `accuracy` が学習データの分類精度、`val_accuracy` がテストデータの分類精度を示す。また、学習結果(損失関数)を図 4.2 に示す。図中の凡例は、`loss` が学習データの損失関数の値、`val_loss` がテストデータの損失関数の値を示す。図 4.1 および図 4.2 より、学習データの学習の精度が上がるにつれ、テストデータの学習の精度が下がっていることが確認された。このことから過学習が起きてしまっていると考えられる。

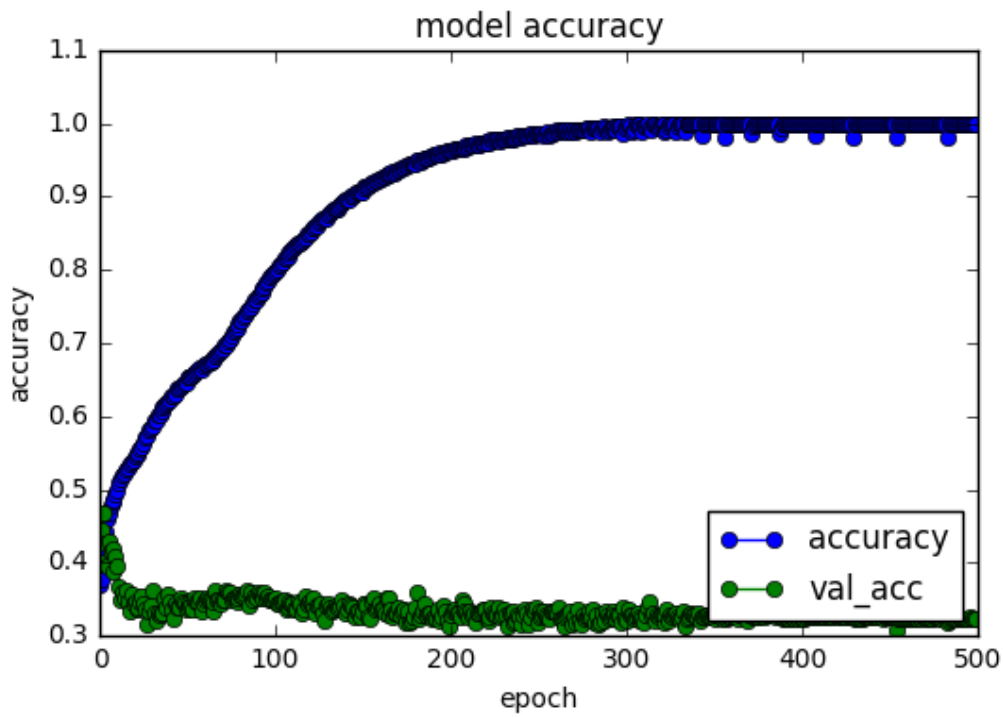


図 4.1 検証 5 の学習結果(accuracy)

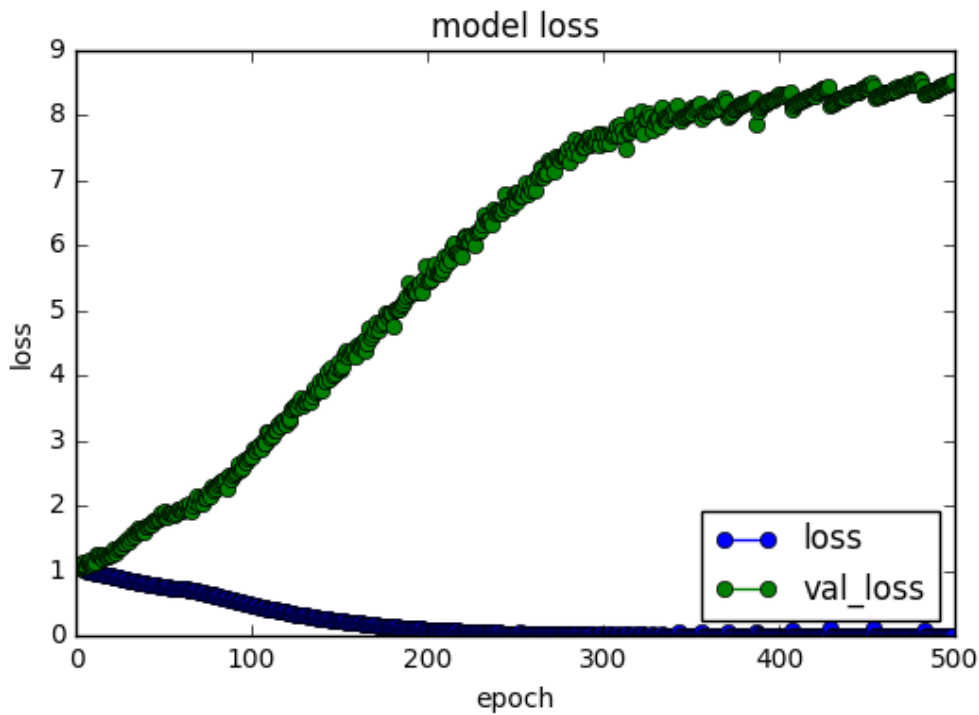


図 4.2 検証 5 の学習結果(loss)

## 4.3.2 検証 6

検証 5 の結果から過学習が起きてしまっていたと考えられたため、Dropout 層を追加して再度検証を行った。学習時の層構造を表 4.5 に示す。

表 4.5 検証 6 におけるネットワーク構造

活性化関数	Sigmoid(最後の層の後のみ Softmax)
最適化アルゴリズム	Adam
層数(ノード数)	6(180,900,400(20%Dropout), 200(20%Dropout),50,3)
トレーニングエポック数	2000
損失関数	categorical_crossentropy

学習結果(分類精度)を図 4.3 に示す。図中の凡例は、accuracy が学習データの分類精度、val\_accuracy がテストデータの分類精度を示す。また、学習結果(損失関数)を図 4.4 に示す。図中の凡例は、loss が学習データの損失関数の値、val\_loss がテストデータの損失関数の値を示す。検証 5 の結果と比較して、Dropout 層を追加したことで学習の進み方が緩やかになったのに対し、テストデータの分類精度が低いままであることに変化はなく精度の向上は見られなかった。

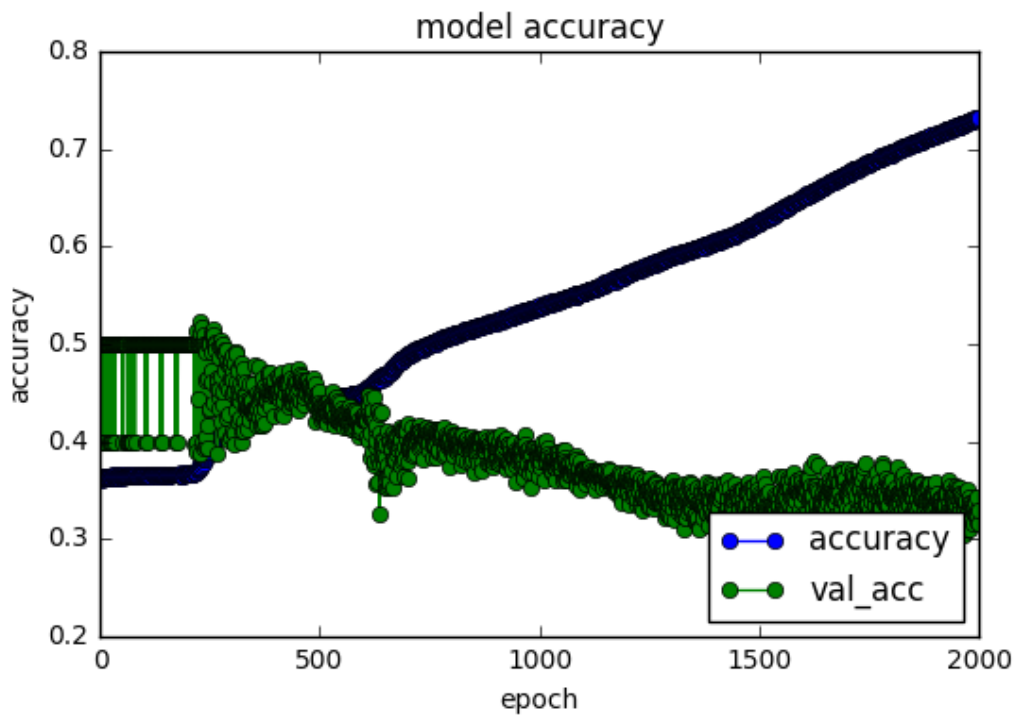


図 4.3 検証 6 の学習結果(accuracy)

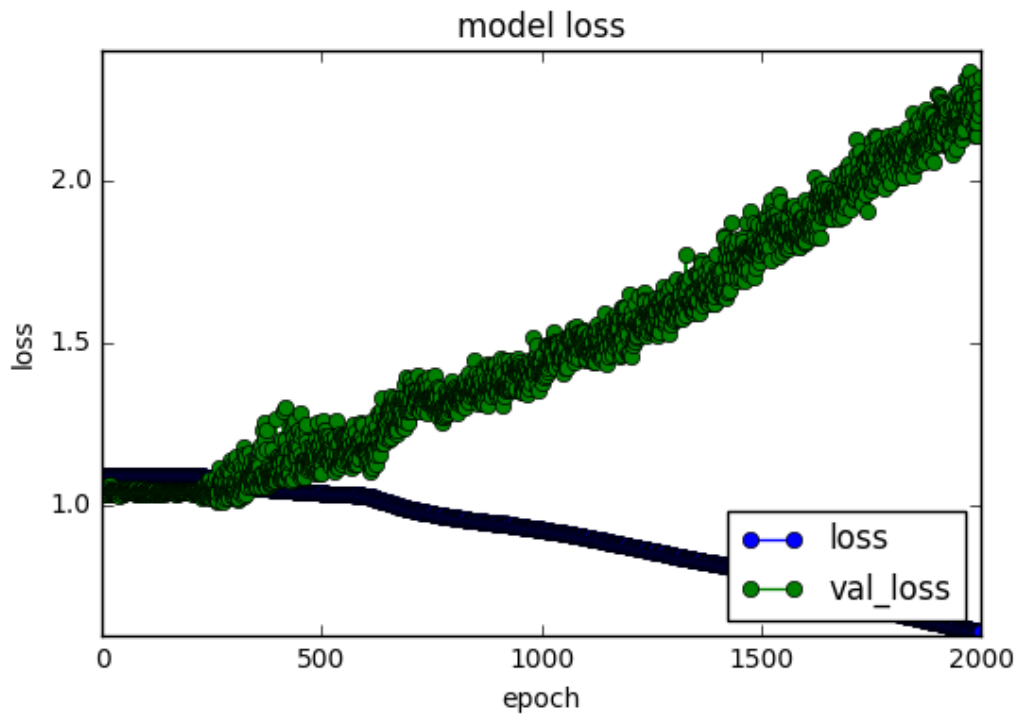


図 4.4 検証 6 の学習結果(loss)

### 4.3.3 検証 7

検証 5 および、検証 6 では、1 秒の窓を 0.05 秒ずつスライドさせてフーリエ変換を施したデータを 1 つずつ使用して解析を行っていた。そのため、データの時間方向の範囲が 1 秒と狭く、精度が上がらない可能性が考えられる。そこで、LSTM(Long short-term memory)を使用することで精度が改善する可能性が考えられる。LSTM とは、RNN(Recurrent Neural Network)を拡張したもので、時系列データに用いられるモデルの一種である。検証 7 では、LSTM を導入することで精度の向上が見られるかどうかの検証を行う。

学習時の層構造を表 4.6 に示す。

表 4.6 検証 7 におけるネットワーク構造

活性化関数	Sigmoid(最後の層の後のみ Softmax)
最適化アルゴリズム	Adam
層数(ノード数)	7(180,180(LSTM 層),180,200,100,50,3)
トレーニングエポック数	2500
損失関数	categorical_crossentropy

学習結果(分類精度)を図 4.5 に示す。図中の凡例は、accuracy が学習データの分類精度、val\_accuracy がテストデータの分類精度を示す。また、学習結果(損失関数)を図 4.6 に示す。図中の凡例は、loss が学習データの損失関数の値、val\_loss がテストデータの損失関数の値を示す。

LSTM 層を追加した場合もほかの場合と同様に過学習が起きてしまいうまく学習が進まず、精度の改善は見られなかった。

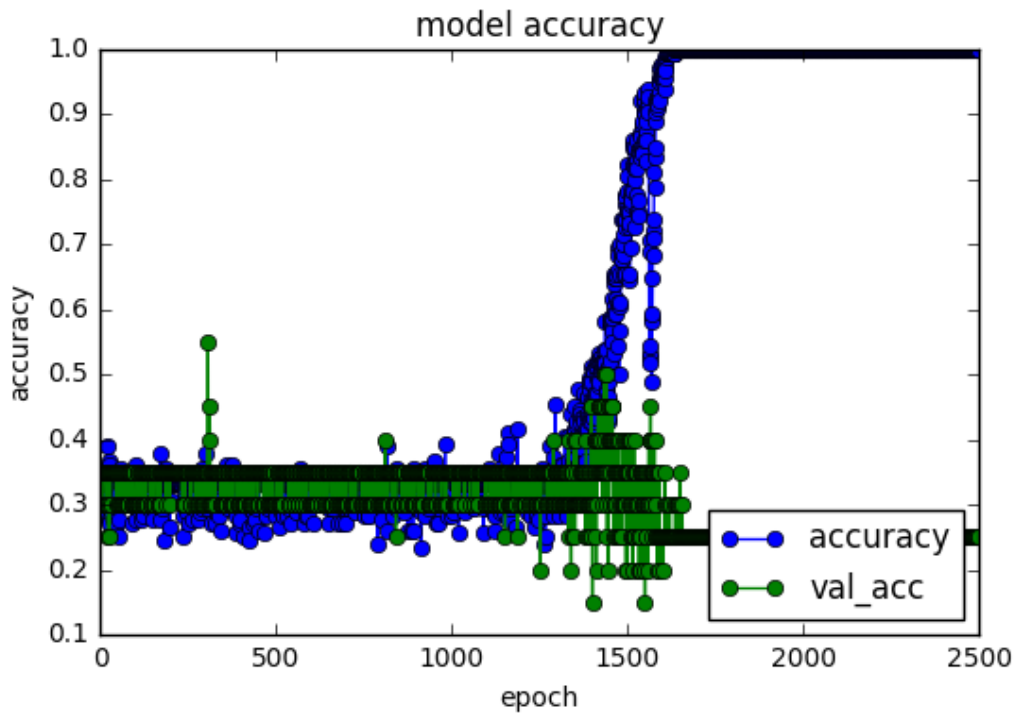


図 4.5 検証 7 の学習結果(accuracy)

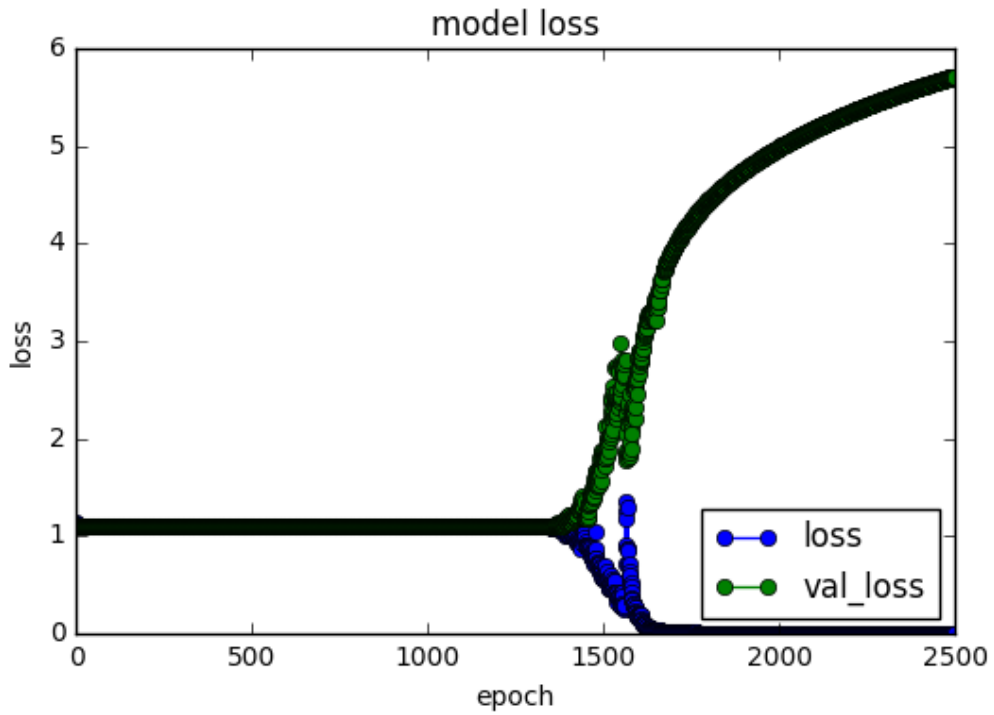


図 4.6 検証 7 の学習結果(loss)

### 4.3.4 検証 8

検証 8 では畳み込み層とプーリング層の利用を試みた。畳み込みと、プーリングを行うことで、データの特徴をより効率的に学習することができる可能性が考えられる。その結果として、各印象評定項目と対応している基礎律動データを抽出しやすくなる可能性が考えられる。畳み込み層とプーリング層を用いることで分類精度の向上が見られるかどうかの検証を行った。

学習時の層構造を表 4.7 および表 4.8 に示す。

表 4.7 検証 8 におけるネットワーク構造

活性化関数	Sigmoid(最後の層の後のみ Softmax)
最適化アルゴリズム	Adam
層数(全結合層のノード数)	9(入力層,畳み込み層,プーリング層, LSTM 層,180,200,100,50,3)
トレーニングエポック数	2500
損失関数	categorical_crossentropy

表 4.8 検証 8 における入力層、畳み込み層、プーリング層、LSTM 層の詳細

入力層	1 × 1180 × 180
畳み込み層	1 × 1131 × 256 (フィルタの長さ:50、ストライド幅:1)
プーリング層	1 × 22 × 256 (プーリングを適応するサイズ:50、 ストライド幅:50)
LSTM 層	1 × 1 × 300

学習結果(分類精度)を図 4.7 に示す。図中の凡例は、accuracy が学習データの分類精度、val\_accuracy がテストデータの分類精度を示す。また、学習結果(損失関数)を図 4.8 に示す。図中の凡例は、loss が学習データの損失関数の値、val\_loss がテストデータの損失関数の値を示す。CNN 層を追加した場合、epoch15 付近で accuracy が 0.5 と十分とは言えないが、ほかの検証に比べ高い精度になった。しかし、その際の損失関数を見るとテストデータの損失関数の値が上昇しており、過学習が起きてしまい学習に失敗している可能性が考えられる。

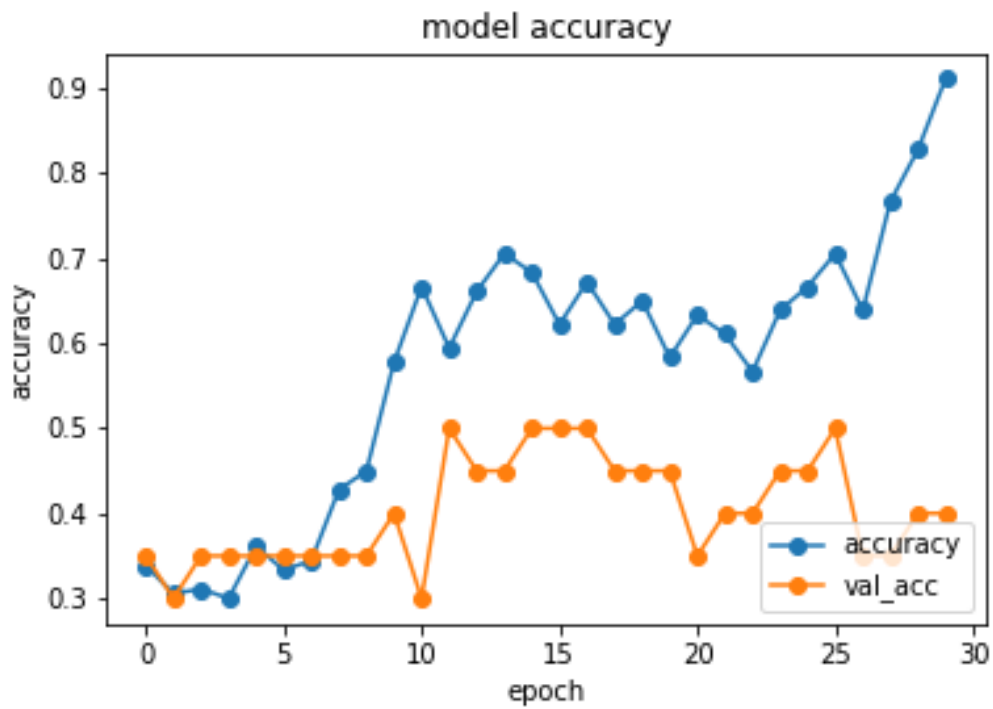


図 4.7 検証 8 の学習結果(accuracy)

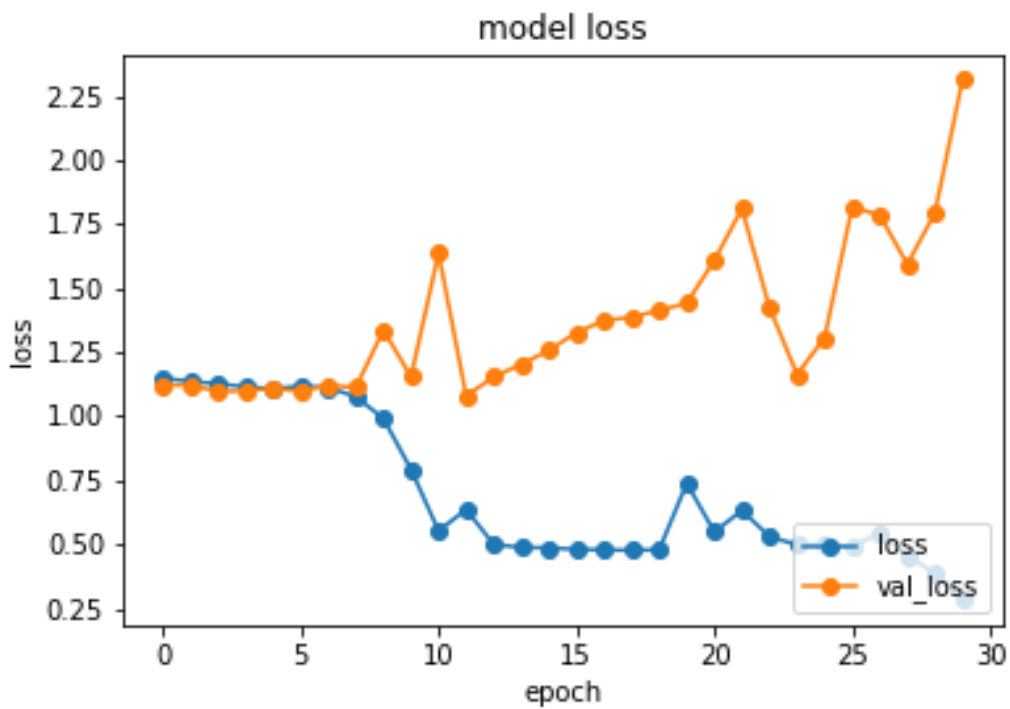


図 4.8 検証 8 の学習結果(loss)



## 4.4 考察とまとめ

4 章ではアンケートの回答結果と楽曲聴取中の基礎律動との間に対応関係があるかどうかの検討を行った。しかし、検証 1 から検証 8 の結果はどれも精度が低く、十分とは言えない結果だった。

SVM に入力するデータの前処理方法を工夫しても大きく精度の変化が見られなかったことから、学習がうまくできていない可能性が考えられる。学習の精度が低いため、不完全な分類器になってしまい、外挿の精度が上がらなかったと考えられる。また、ディープラーニングの学習結果を見てもテストデータの分類精度が低い状態から過学習が起きてしまっており、学習に失敗している。そのため、より高い精度で分類するためには機械学習を行う際のハイパーパラメータの調整が必要であると考えられる。

また、正解データ（アンケート）が正確ではないため分類精度が低下してしまっている可能性などが考えられる。アンケートは 1 音源に対して 1 つだけ取っているのに対し、FFT を施した後の基礎律動データは 1 音源当たり 1180 データと、1 データあたりの時間幅に大きな差がある。そのため、アンケートの回答結果に十分な分解能がなく、被験者の楽曲に対する印象の変動に対応できず、精度が低下してしまった可能性が考えられる。

## 第5章 聴取楽曲及びジャンルと基礎律動との対応

### に関する解析

4章では、教師あり学習を用いて、アンケート結果を正解データとして学習及び分類を試みた。その結果、十分な分類精度が得られず、正解データ（アンケート）が正確でない可能性が示唆された。そこで5章では、アンケート結果を用いずに教師なし学習であるオートエンコーダを用いて解析を行った。なお、第5章の内容は自著文献[16]の内容を含む。

#### 5.1 オートエンコーダ

オートエンコーダはニューラルネットワークを用いた手法である。ニューラルネットワークの学習時に入力と出力が同じになるようにネットワークを構成し学習を行う。ネットワークを構成する際に、中間ノード数を入力次元数よりも小さくすることで、オートエンコーダは入力データを少ない次元で復元できるように入力データの特徴表現を生成する。このように生成された特徴表現には、入力データの特徴が反映されている。オートエンコーダはこのような特徴を持つため、次元縮退のためのアルゴリズムとしてよく用いられている。

#### 5.2 オートエンコーダを用いた学習

本研究では3種類のオートエンコーダを作成し、検証を行った。入力データとして3.7.1で算出したパワースペクトルデータにL2正規化を施したデータを使用した。学習時にテストデータとして音源番号1-20を聴取している時の基礎律動データをテストデータとして使用し、残りの音源を聴取している時の基礎律動データを学習データとして使用した。特徴表現の算出は学習を行った中で最もテストデータの損失関数の値が小さかったものを使用した。また、学習時には各チャンネル個別に学習を行った。学習の精度を評価する指標として損失関数の他に平均絶対誤差率(MAPE)を使用した。今回は、学習データおよびテストデータのMAPEの値が20-25の範囲で学習が収束することを目標に学習を行った。オートエンコーダによる学習及び特徴表現の算出の概要図を図5.1に示す。オートエンコーダの実装にはKeras[12]を使用した。Kerasのバックエンドは、Tensorflow[13]を使用した。

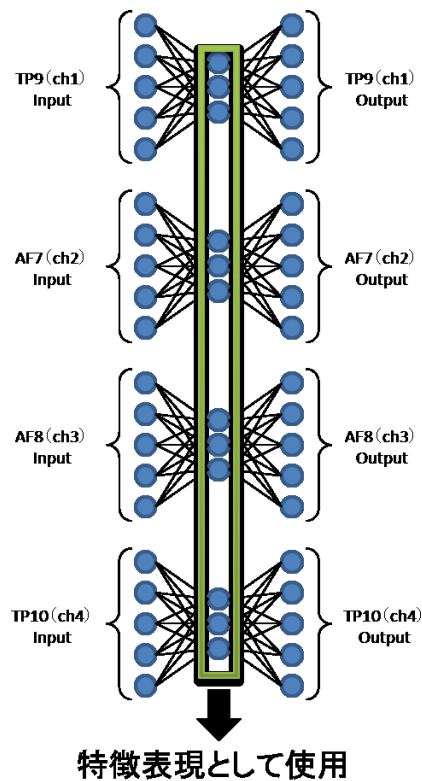


図 5.1 オートエンコーダ学習の概要図

## 5.2.1 ネットワーク 1

ネットワーク 1 では、3.7.1 で算出したパワースペクトルの中から、1-30[Hz]のものに L2 正規化を適応したデータを用いて学習を行った。実際に使用した層構造を表 5.1 に示す。

表 5.1 ネットワーク 1 の層構造

活性化関数	ReLU
最適化アルゴリズム	Adam
層数(全結合層のノード数)	11(30,4096,1024,512,64,20,64,512,1024,4096,30)
トレーニングエポック数	300
損失関数	MSE (Mean Squared Error)

学習結果として、学習データの MSE のグラフを図 5.2 から図 5.5 に示す。テストデータの MSE を示したグラフを図 5.6 から図 5.9 に示す。また、学習データの MAPE のグラフを図 5.10 から図 5.13 に、テストデータの MAPE のグラフを図 5.14 から図 5.17 に示す。

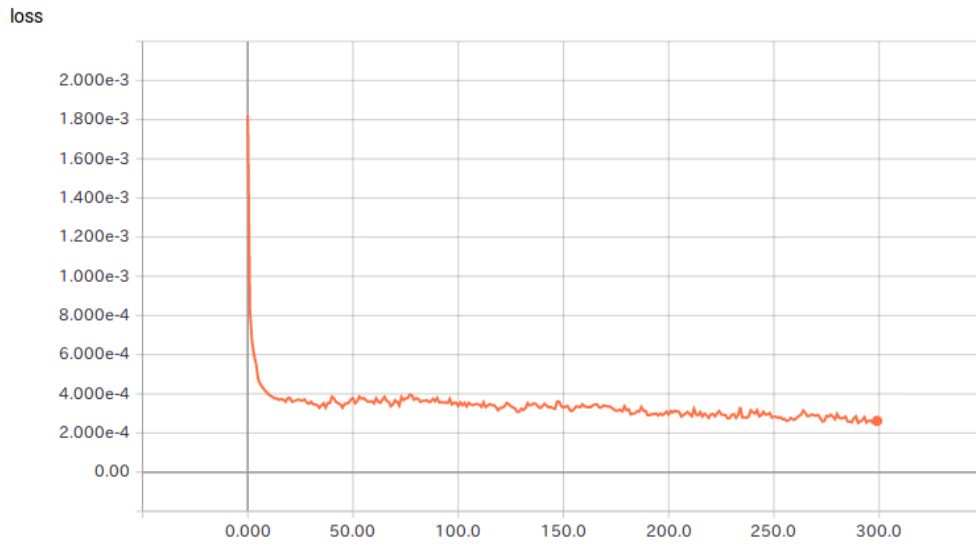


図 5.2 ch1 における学習データの MSE(ネットワーク 1)

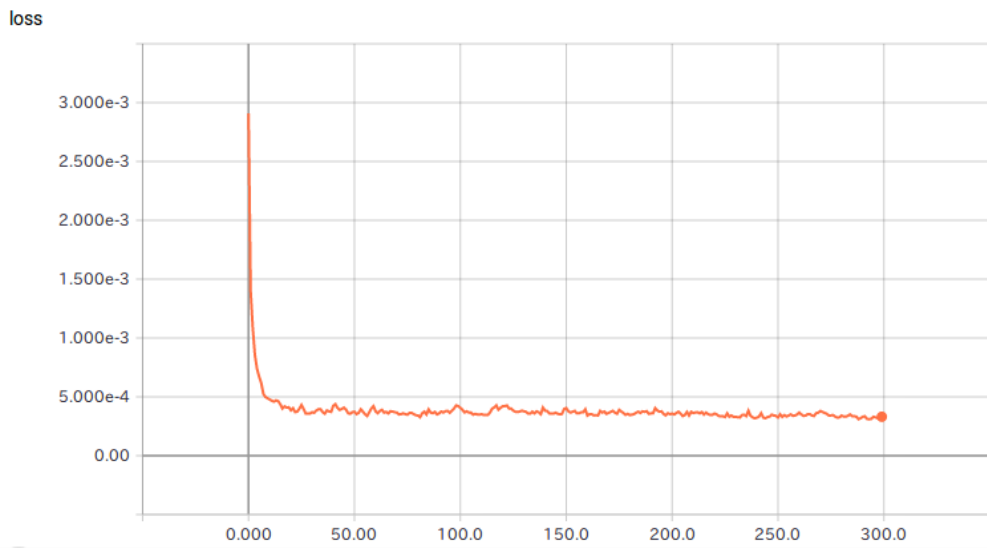


図 5.3 ch2 における学習データの MSE(ネットワーク 1)

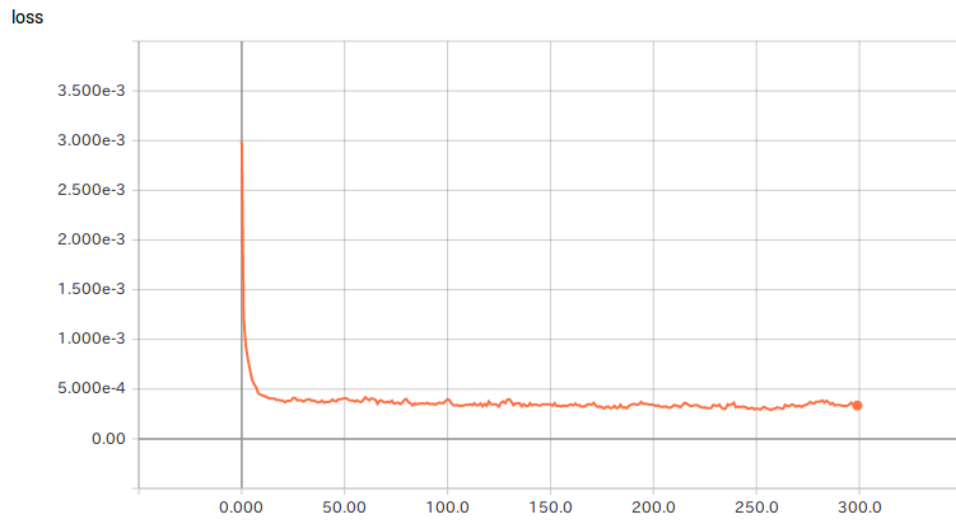


図 5.4 ch3 における学習データの MSE(ネットワーク 1)

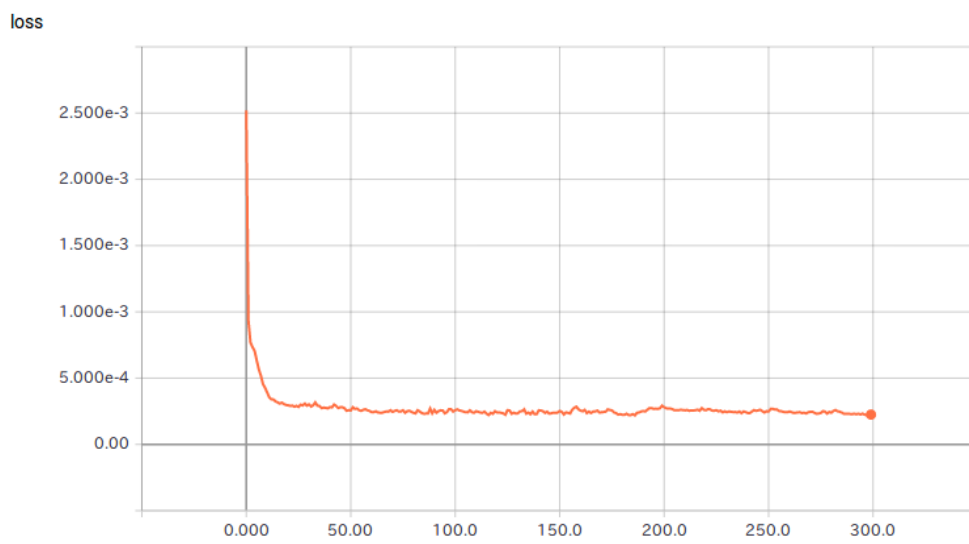


図 5.5 ch4 における学習データの MSE(ネットワーク 1)

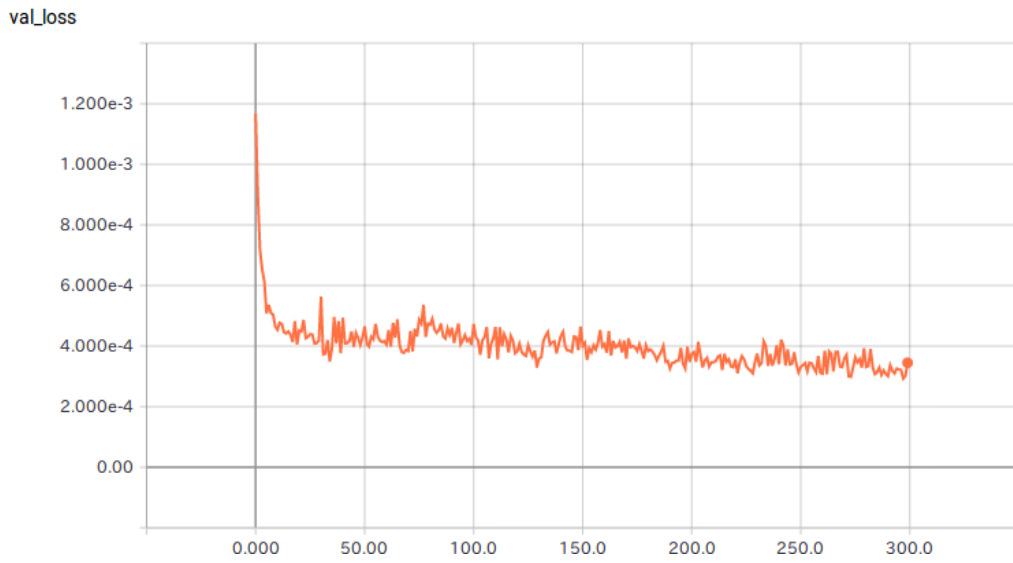


図 5.6 ch1 におけるテストデータの MSE(ネットワーク 1)

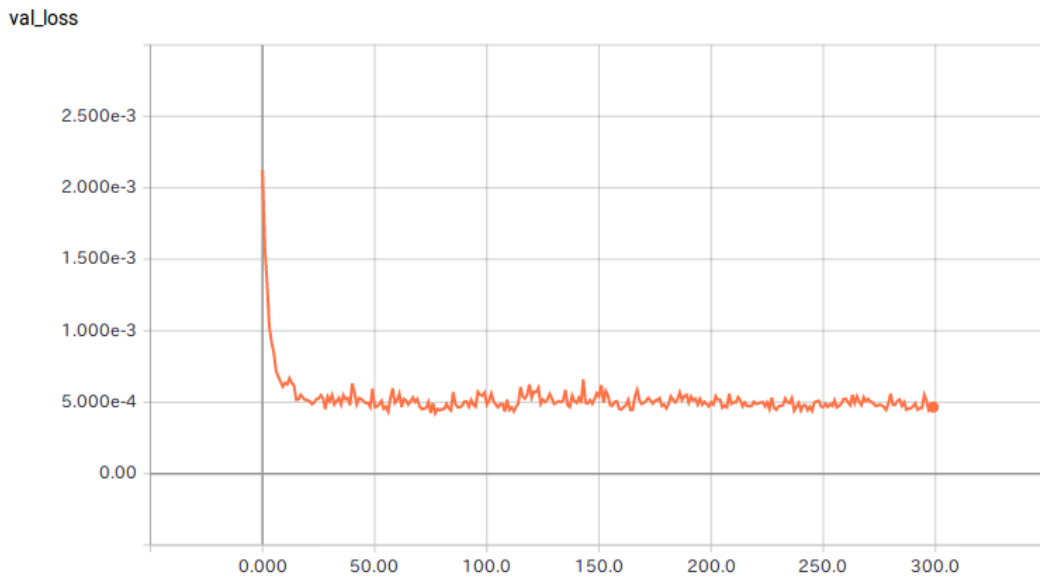


図 5.7 ch2 におけるテストデータの MSE(ネットワーク 1)

val\_loss

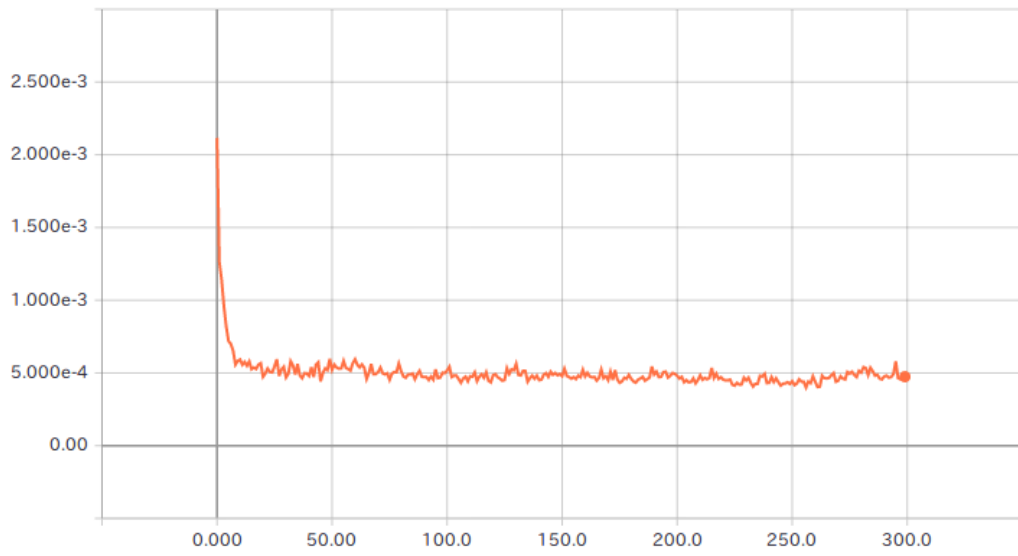


図 5.8 ch3 におけるテストデータの MSE(ネットワーク 1)

val\_loss

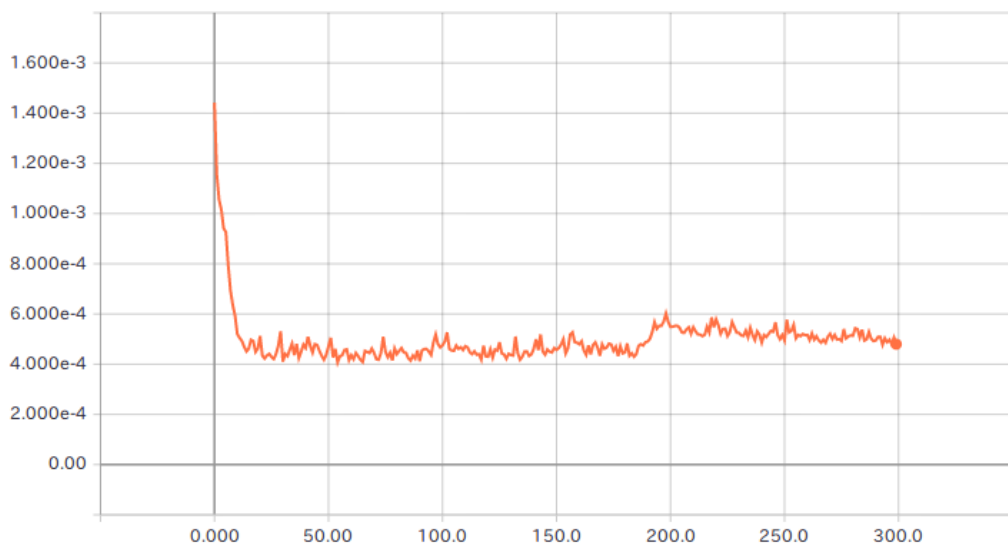


図 5.9 ch4 におけるテストデータの MSE(ネットワーク 1)

mean\_absolute\_percentage\_error

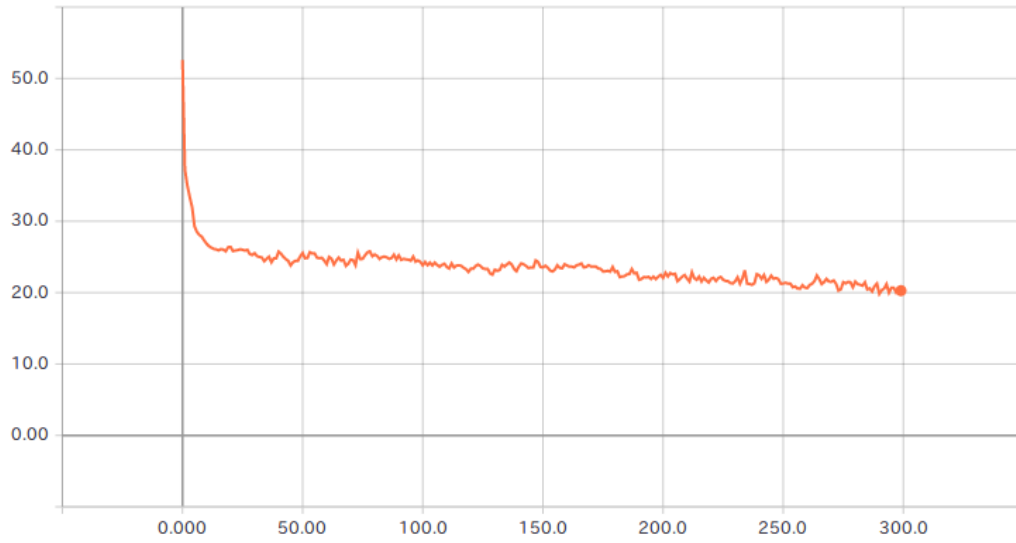


図 5.10 ch1 における学習データの MAPE(ネットワーク 1)

mean\_absolute\_percentage\_error

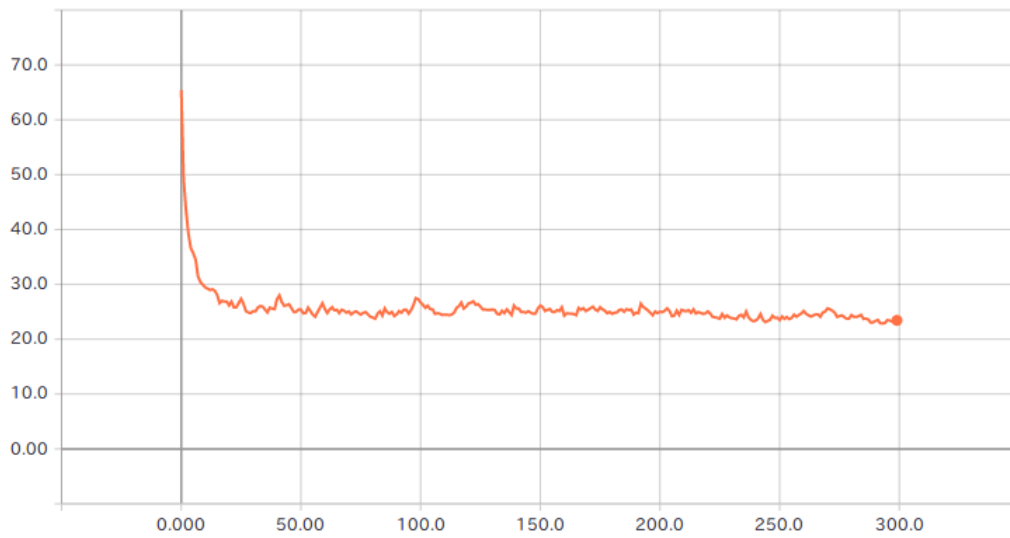


図 5.11 ch2 における学習データの MAPE(ネットワーク 1)



mean\_absolute\_percentage\_error

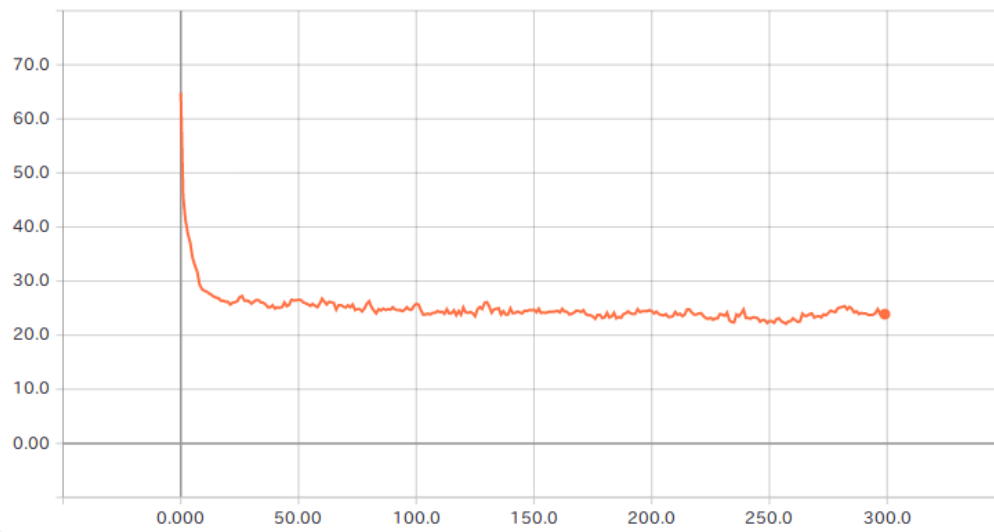


図 5.12 ch3 における学習データの MAPE(ネットワーク 1)

mean\_absolute\_percentage\_error

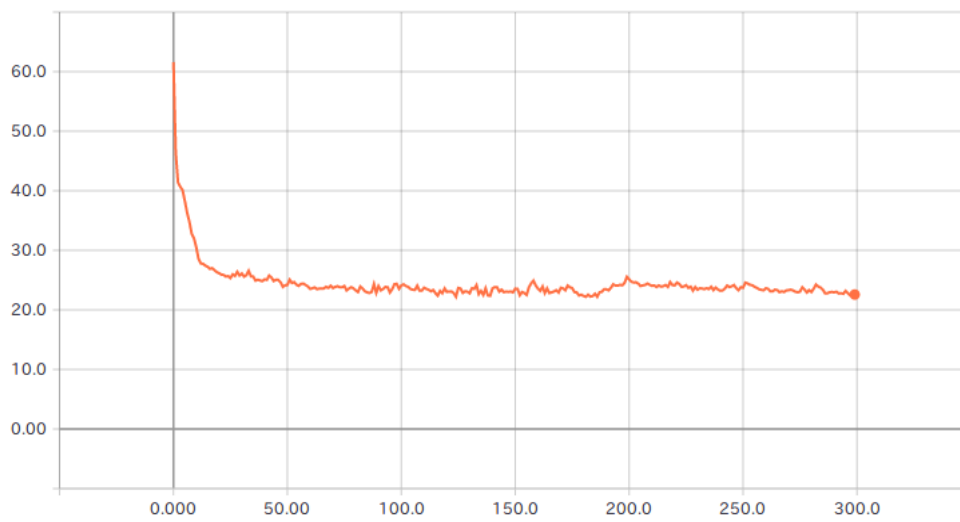


図 5.13 ch4 における学習データの MAPE(ネットワーク 1)

val\_mean\_absolute\_percentage\_error

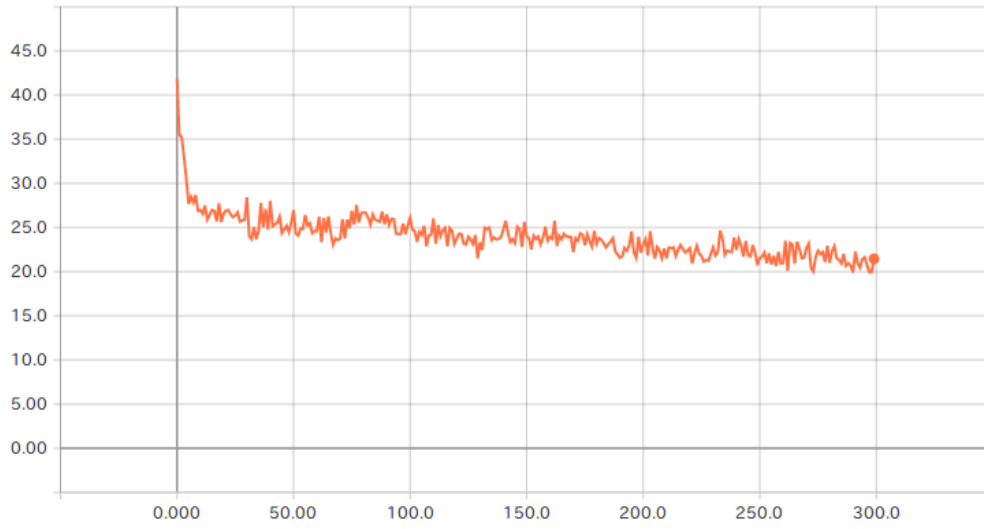


図 5.14 ch1 におけるテストデータの MAPE(ネットワーク 1)

val\_mean\_absolute\_percentage\_error

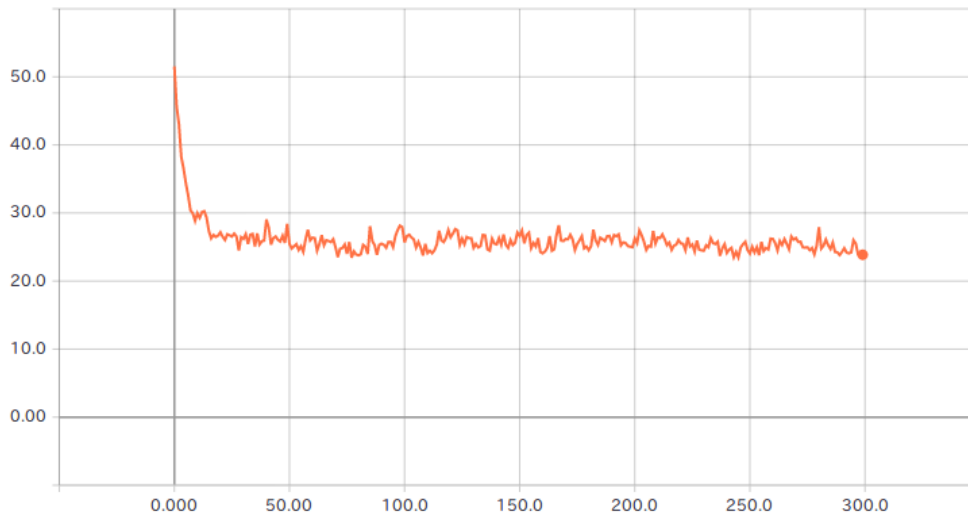


図 5.15 ch2 におけるテストデータの MAPE(ネットワーク 1)

val\_mean\_absolute\_percentage\_error

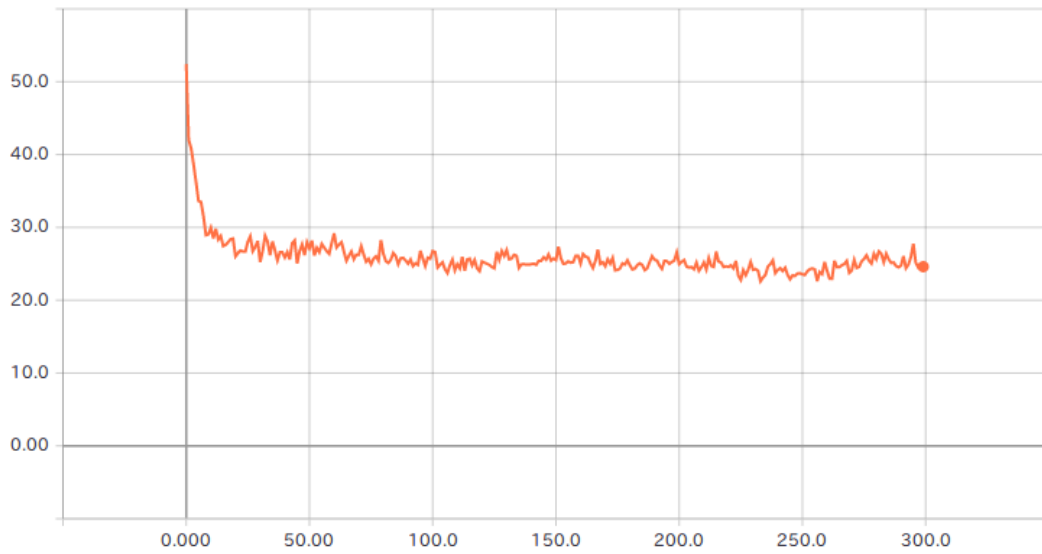


図 5.16 ch3 におけるテストデータの MAPE(ネットワーク 1)

val\_mean\_absolute\_percentage\_error

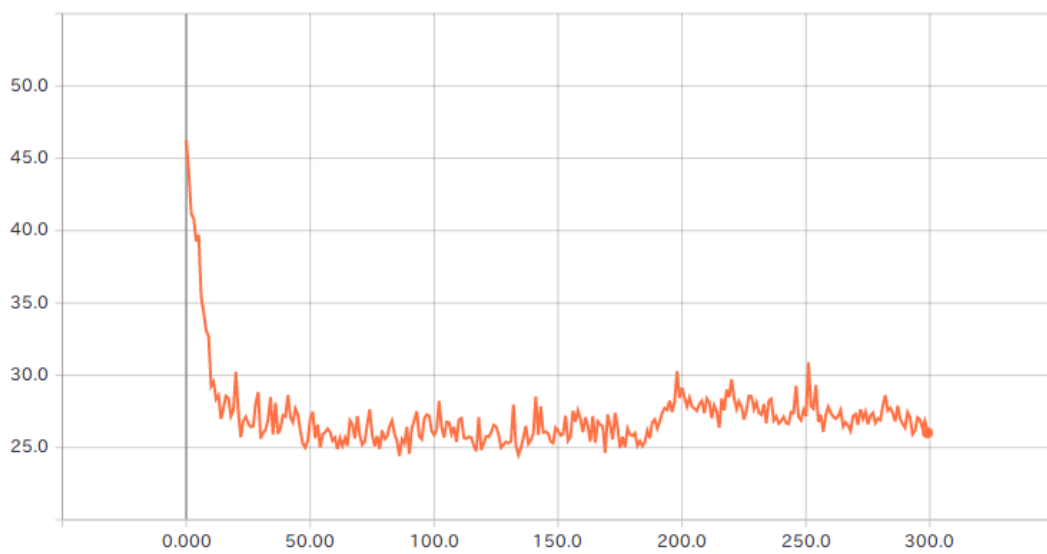


図 5.17 ch4 におけるテストデータの MAPE(ネットワーク 1)

図 5.2 から図 5.17 の結果から学習が収束していることが確認できる。MAPE の値を確認すると、すべてのチャンネルで 20-25 付近の精度で学習が収束していることが確認できる。よって、ch1 から ch4 の学習が成功しているといえる。

## 5.2.2 ネットワーク 2

ネットワーク 2 では、3.7.1 で算出したパワースペクトルの中から、1-45[Hz]すべてに対して L2 正規化を適応したデータを用いて学習を行った。実際に使用した層構造を表 5.2 に示す。

表 5.2 ネットワーク 2 の層構造

活性化関数	ReLU
最適化アルゴリズム	Adam
層数(全結合層のノード数)	11(45,8092,2048,1024,128,35,128,1024,2048,8092,45)
トレーニングエポック数	500
損失関数	MSE (Mean Squared Error)

学習結果として、学習データの MSE のグラフを図 5.18 から図 5.21 に示す。テストデータの MSE を示したグラフを図 5.22 から図 5.25 に示す。また、学習データの MAPE のグラフを図 5.26 から図 5.29 に、テストデータの MAPE のグラフを図 5.30 から図 5.33 に示す。

loss

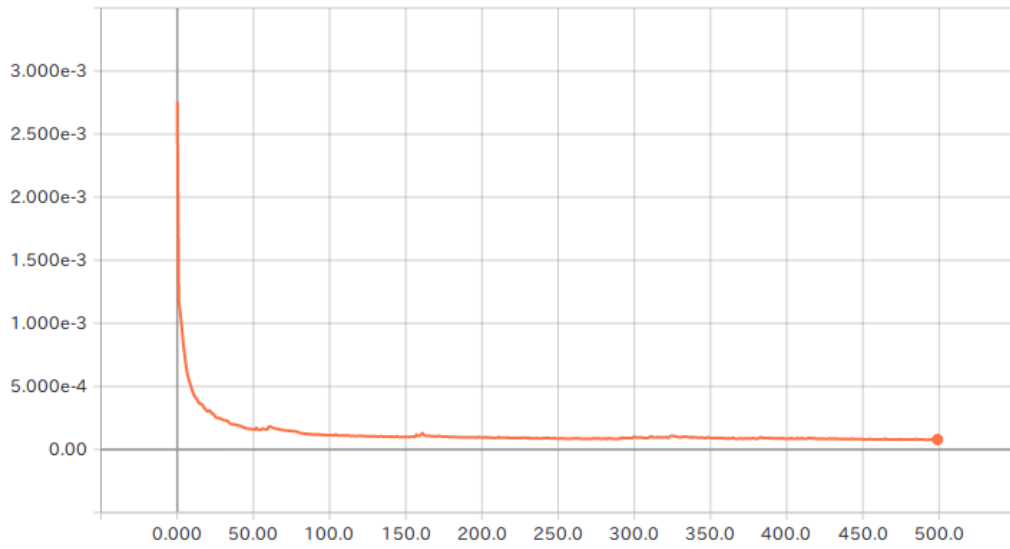


図 5.18 ch1 における学習データの MSE(ネットワーク 2)

loss

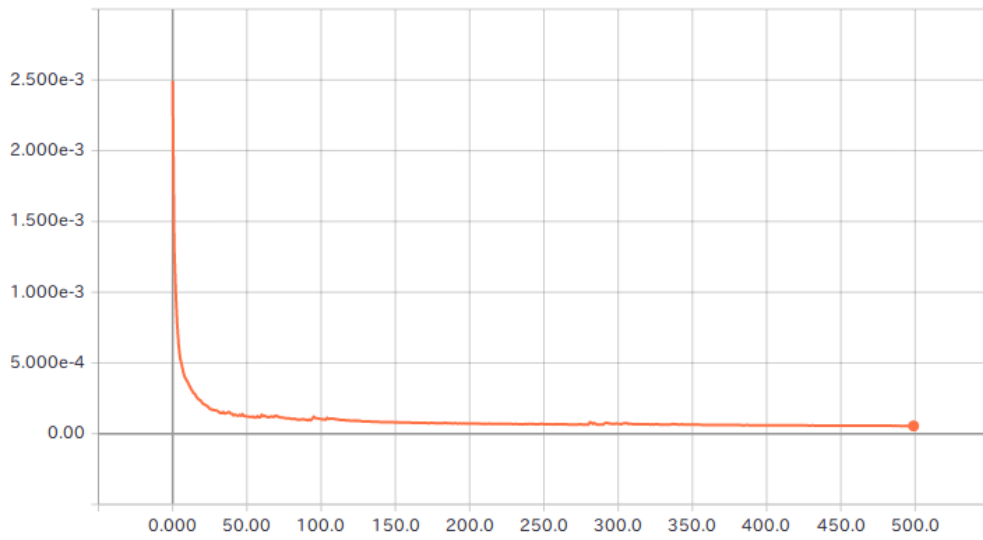


図 5.19 ch2 における学習データの MSE(ネットワーク 2)

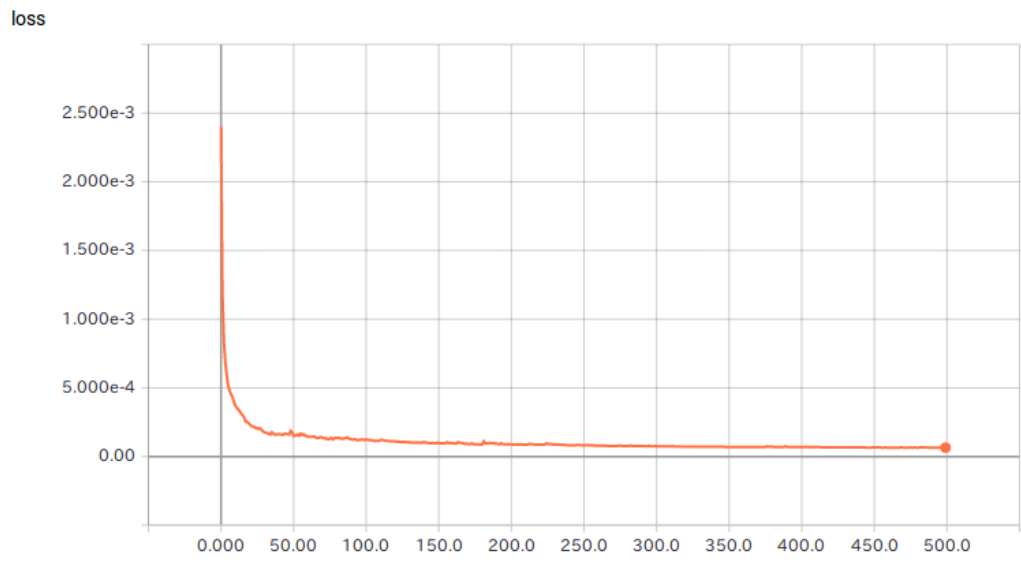


図 5.20 ch3 における学習データの MSE(ネットワーク 2)

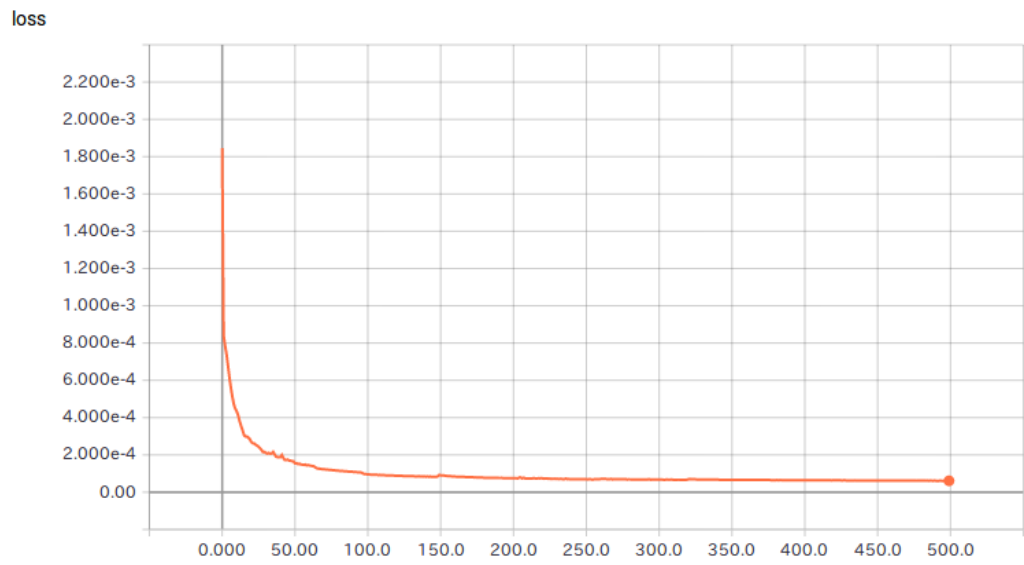


図 5.21 ch4 における学習データの MSE(ネットワーク 2)

val\_loss

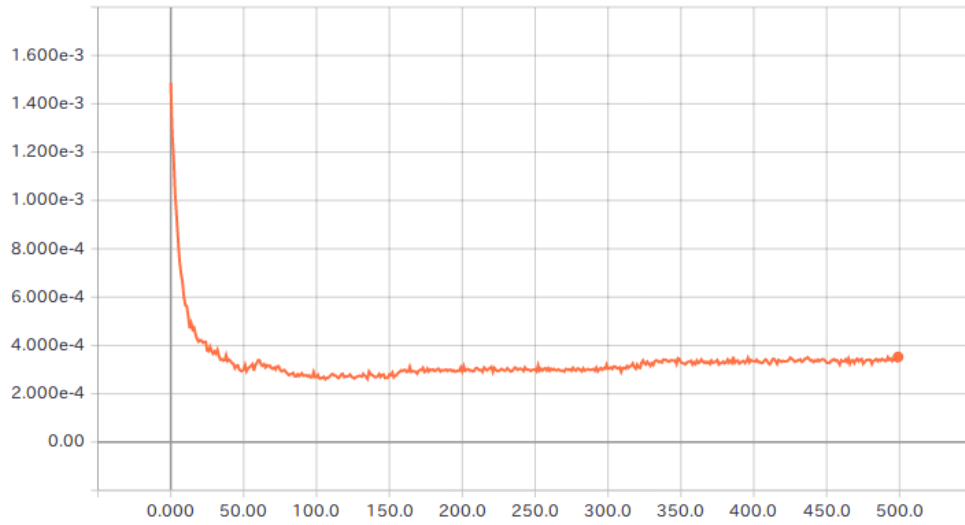


図 5.22 ch1 におけるテストデータの MSE(ネットワーク 2)

val\_loss

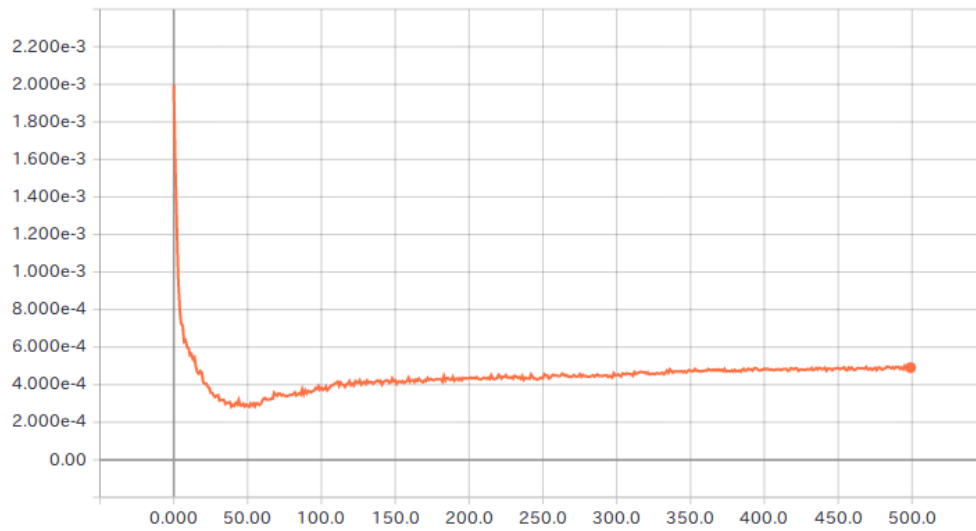


図 5.23 ch2 におけるテストデータの MSE(ネットワーク 2)

val\_loss

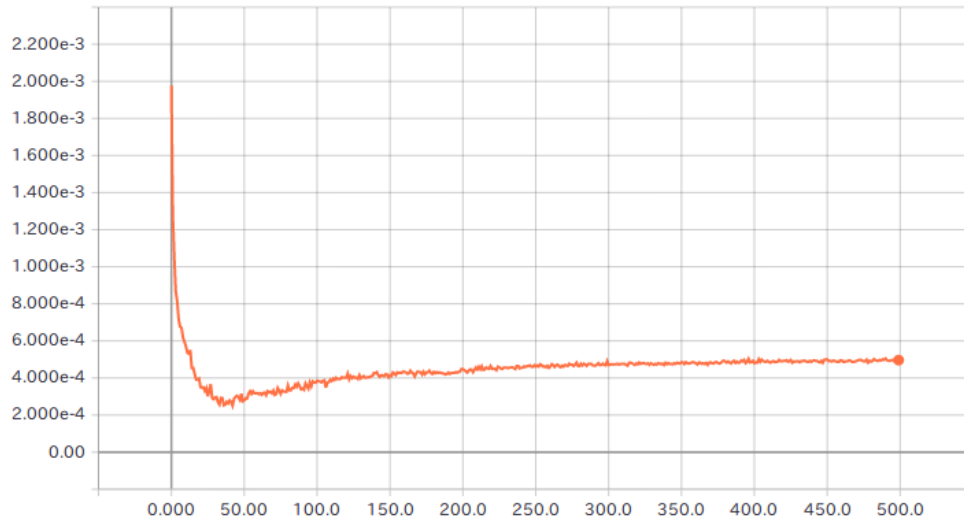


図 5.24 ch3 におけるテストデータの MSE(ネットワーク 2)

val\_loss

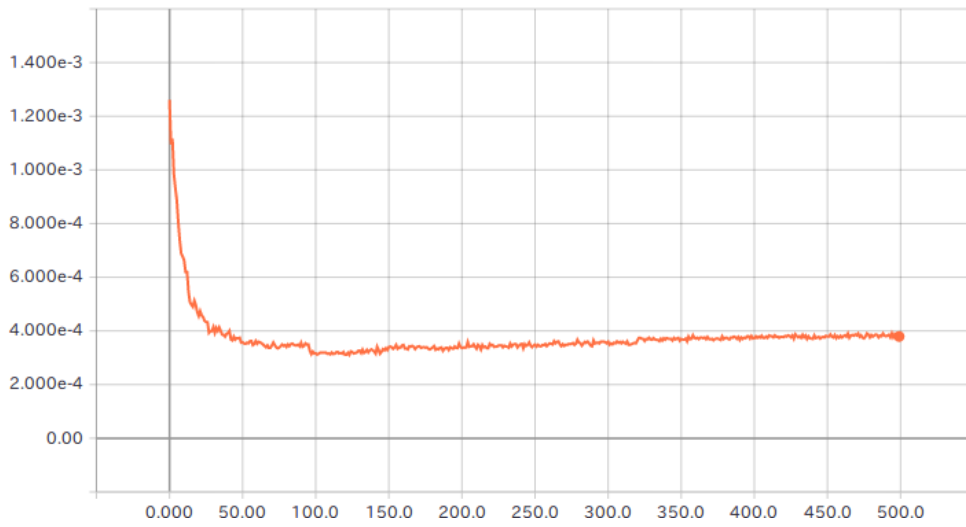


図 5.25 ch4 におけるテストデータの MSE(ネットワーク 2)



mean\_absolute\_percentage\_error

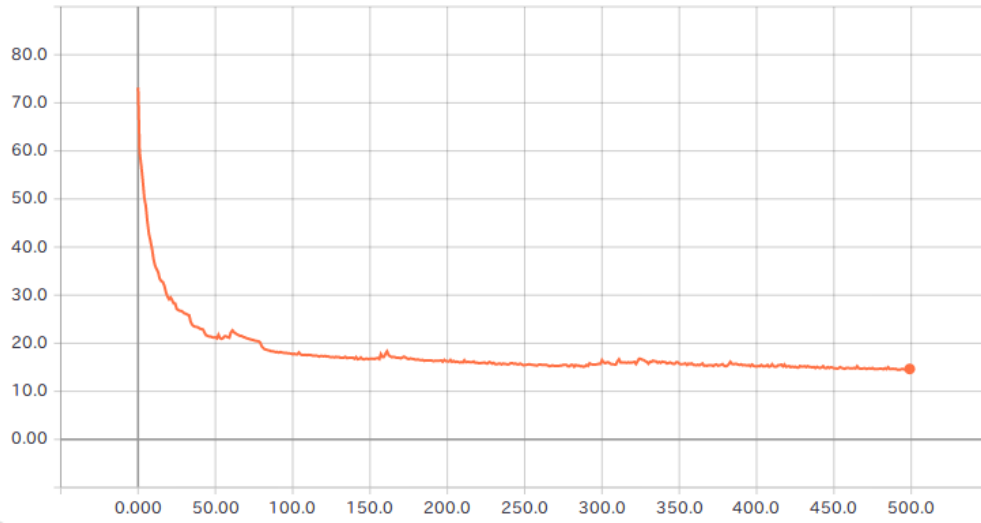


図 5.26 ch1 における学習データの MAPE(ネットワーク 2)

mean\_absolute\_percentage\_error

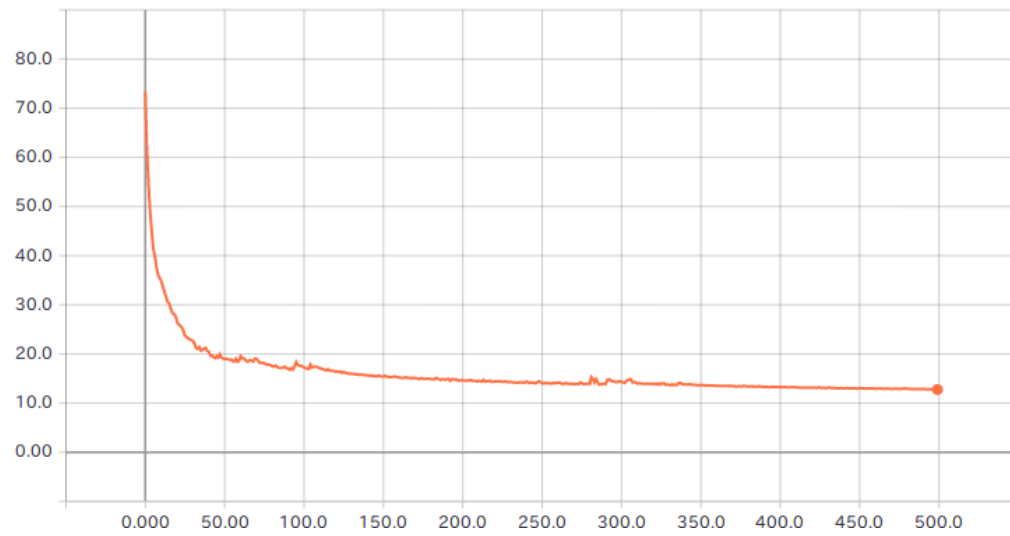


図 5.27 ch2 における学習データの MAPE(ネットワーク 2)

mean\_absolute\_percentage\_error

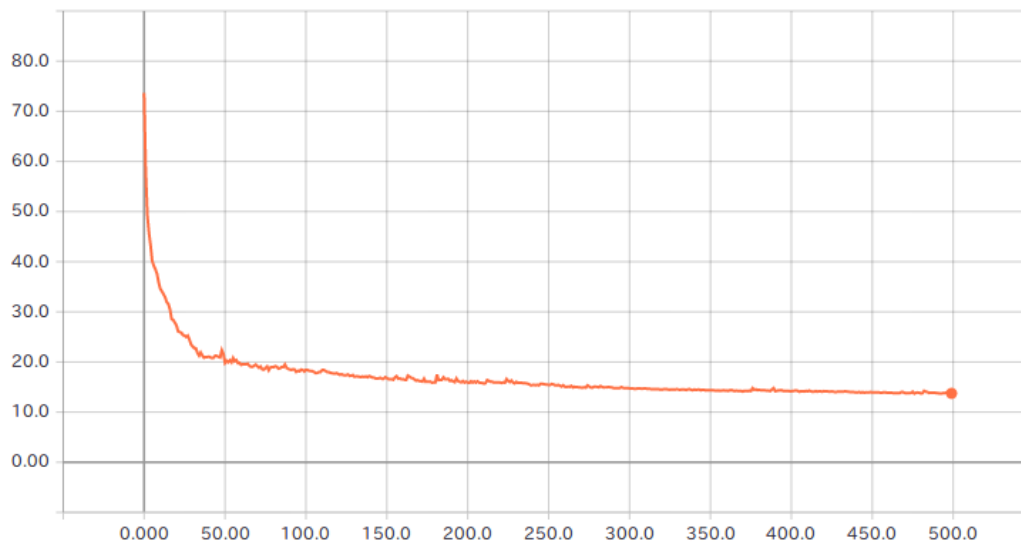


図 5.28 ch3 における学習データの MAPE(ネットワーク 2)

mean\_absolute\_percentage\_error

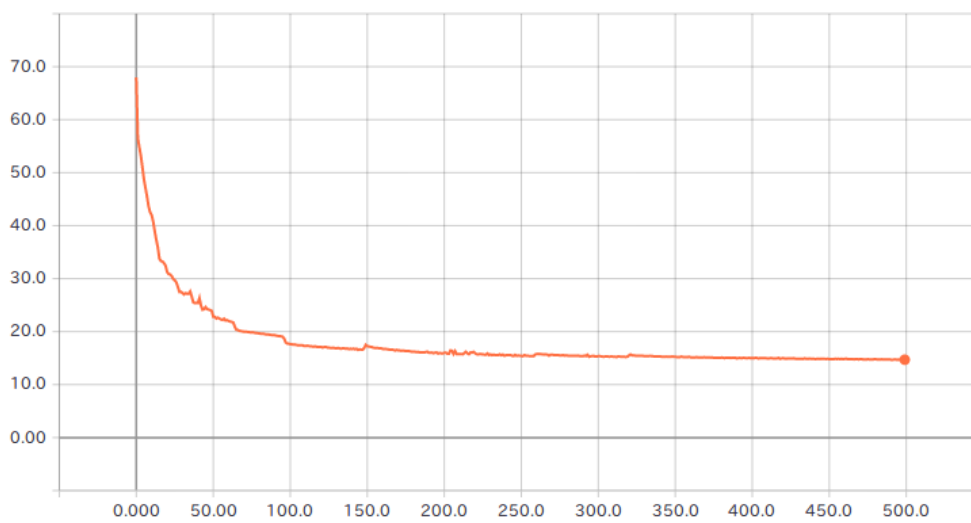


図 5.29 ch4 における学習データの MAPE(ネットワーク 2)

val\_mean\_absolute\_percentage\_error

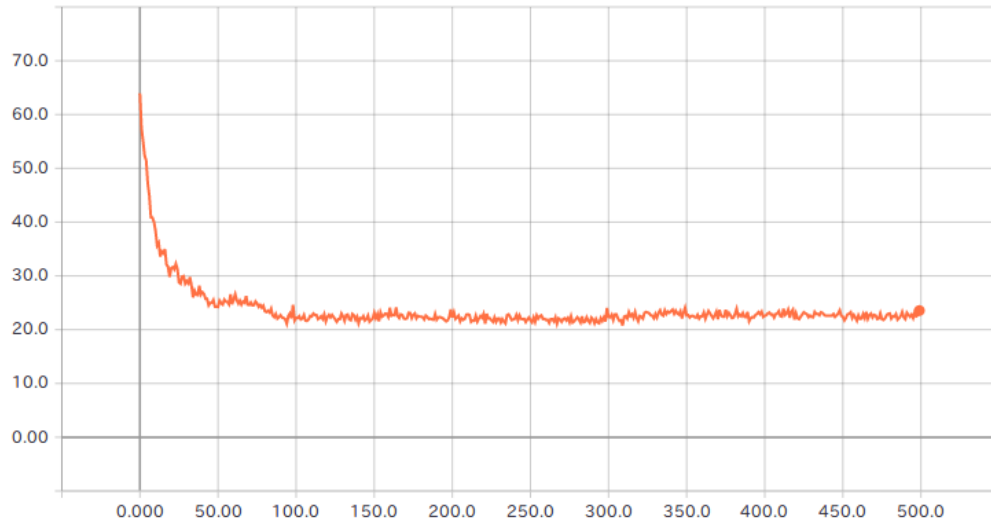


図 5.30 ch1 におけるテストデータの MAPE(ネットワーク 2)

val\_mean\_absolute\_percentage\_error

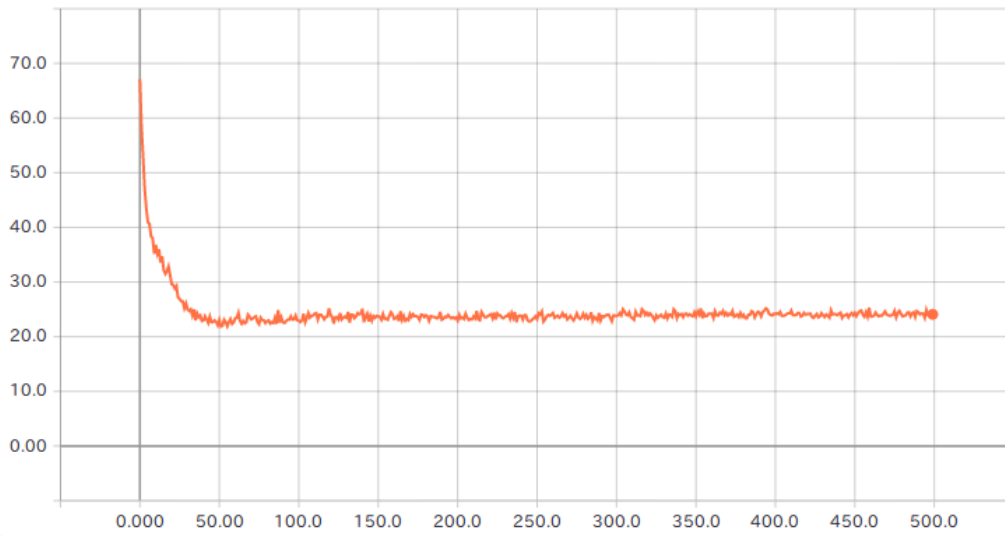


図 5.31 ch2 におけるテストデータの MAPE(ネットワーク 2)

val\_mean\_absolute\_percentage\_error

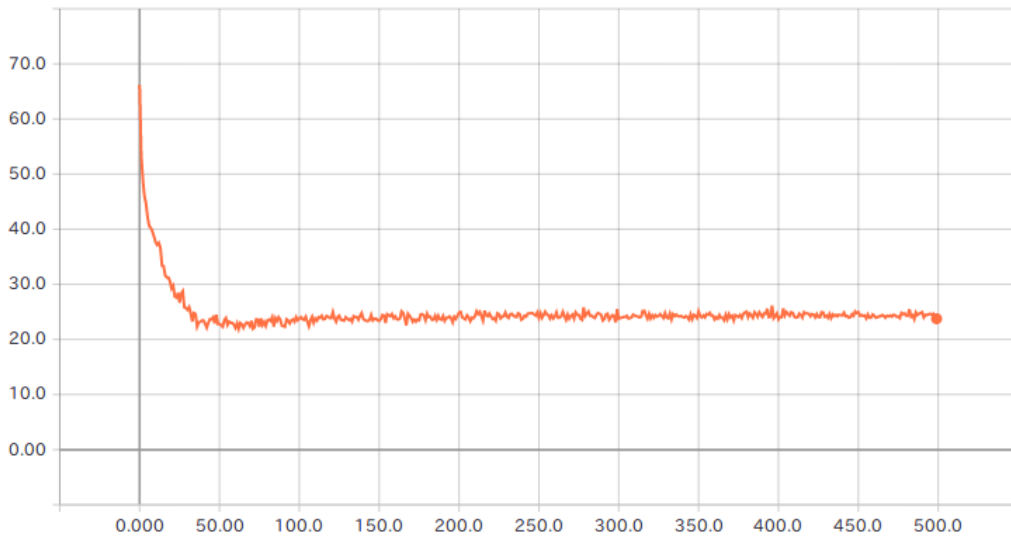


図 5.32 ch3 におけるテストデータの MAPE(ネットワーク 2)

val\_mean\_absolute\_percentage\_error

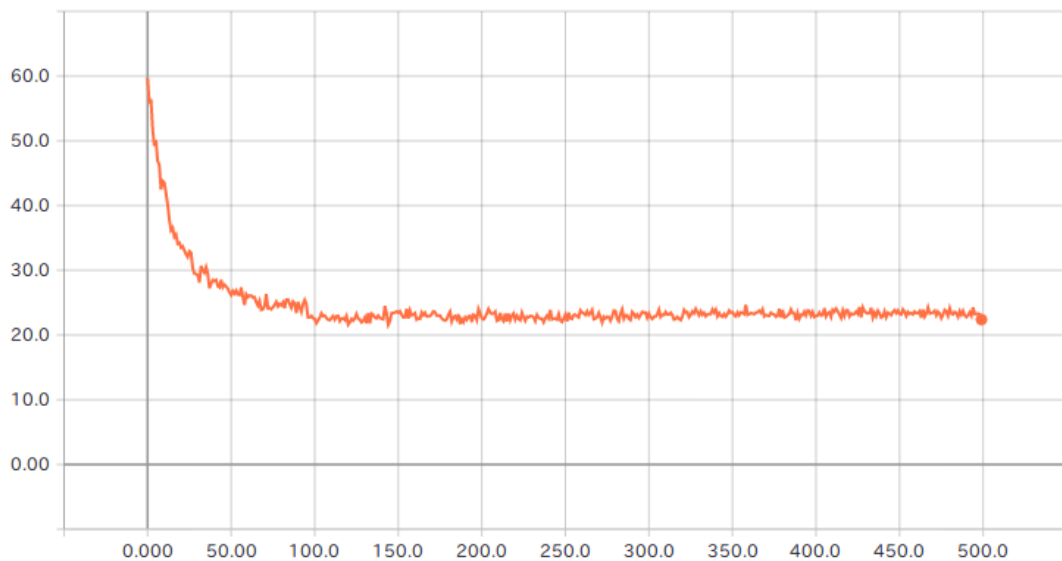


図 5.33 ch4 におけるテストデータの MAPE(ネットワーク 2)

図 5.22 および、図 5.25 より、ch1 と ch4 では、エポック数 100 付近から過学習が起きてしまっていることが確認できる。しかし、図 5.30、図 5.33 を見ると、エポック数 100 付近での MAPE の値は 20-25 であり、学習が成功しているといえる。同様に、ch2 と ch3 では、図 5.23 および、図 5.24 よりエポック数 50 付近から過学習が起きてしまっていることが確認できる。しかし、図 5.31、図 5.32 を確認すると、エポック数 50 付近での MAPE の値が 20-25 であり、学習が成功していると考えられる。これ

らより、全チャンネルとも学習が成功したといえる。

### 5.2.3 ネットワーク 3

ネットワーク 3 では、時間方向の変動を考慮して学習を行うために LSTM のみを使用してオートエンコーダを作成した。3.7.1 で算出したパワースペクトルの中から 1-45[Hz]すべてに対して L2 正規化を適応したデータを用いて学習を行った。実際に使用した層構造を表 5.3 に示す。

表 5.3 ネットワーク 3 の層構造

活性化関数	Tanh
最適化アルゴリズム	RMSprop
層数(LSTM 層のノード数)	5(45,512,35,512,45)
トレーニングエポック数	5000
損失関数	MSE (Mean Squared Error)

学習結果として、学習データの MSE のグラフを図 5.34 から図 5.37 に示す。テストデータの MSE を示したグラフを図 5.38 から図 5.41 に示す。また、学習データの MAPE のグラフを図 5.42 から図 5.45 に、テストデータの MAPE のグラフを図 5.46 から図 5.49 に示す。

loss

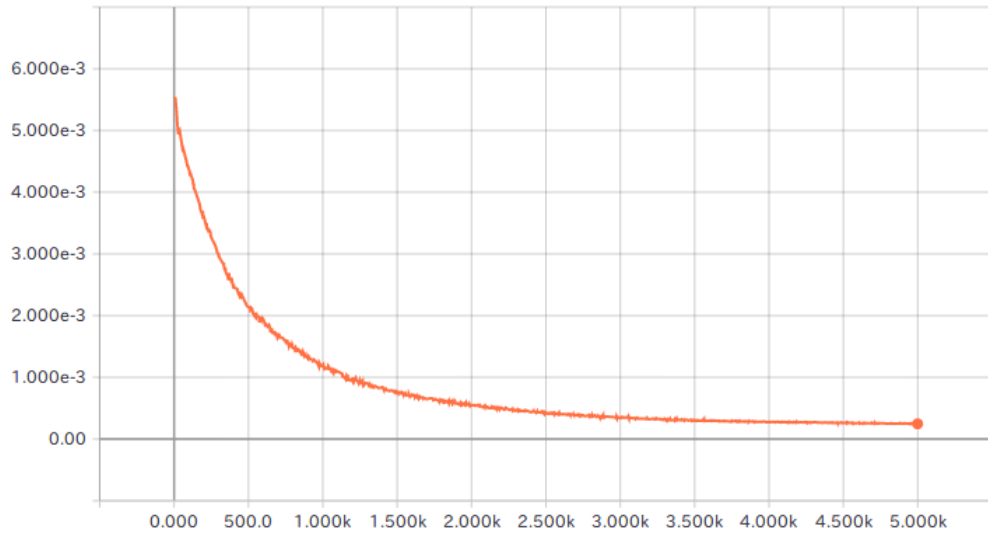


図 5.34 ch1 における学習データの MSE(ネットワーク 3)

loss

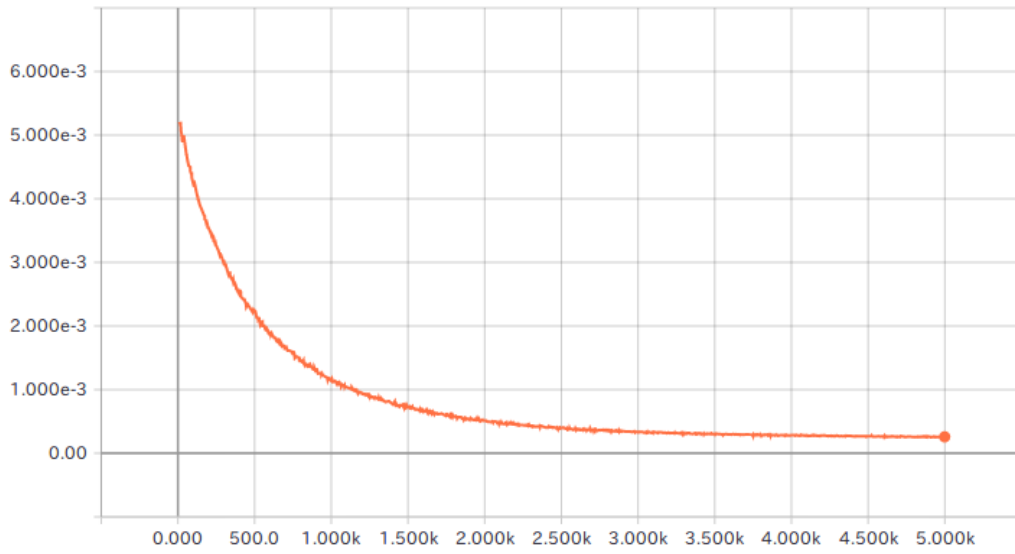


図 5.35 ch2 における学習データの MSE(ネットワーク 3)

loss

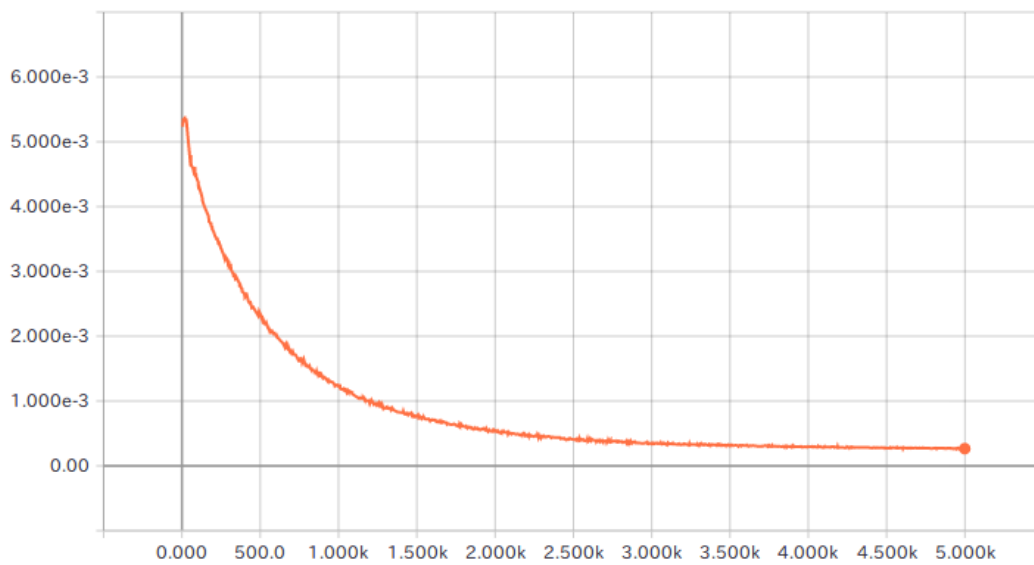


図 5.36 ch3 における学習データの MSE(ネットワーク 3)

loss

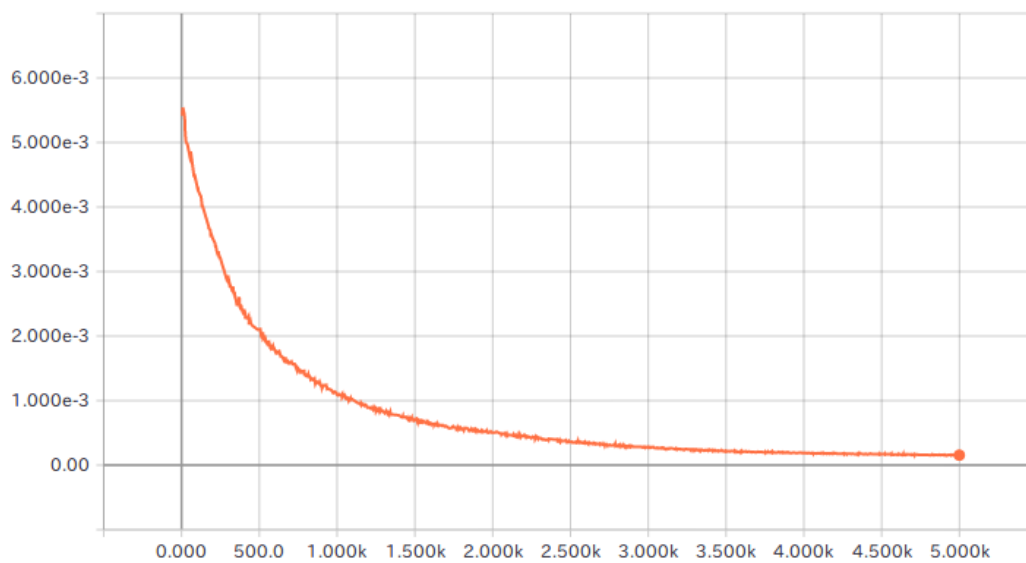


図 5.37 ch4 における学習データの MSE(ネットワーク 3)

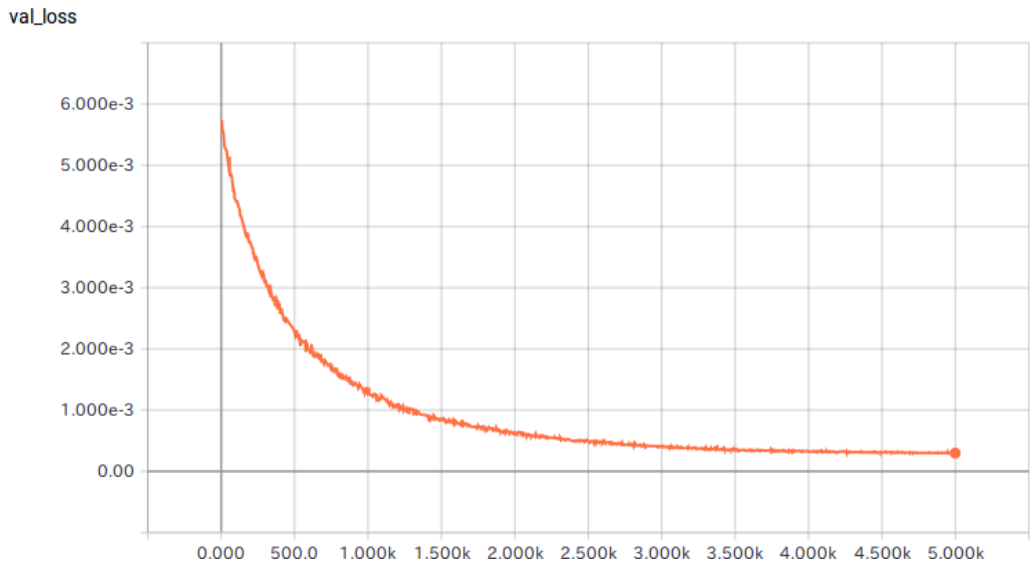


図 5.38 ch1 におけるテストデータの MSE(ネットワーク 3)

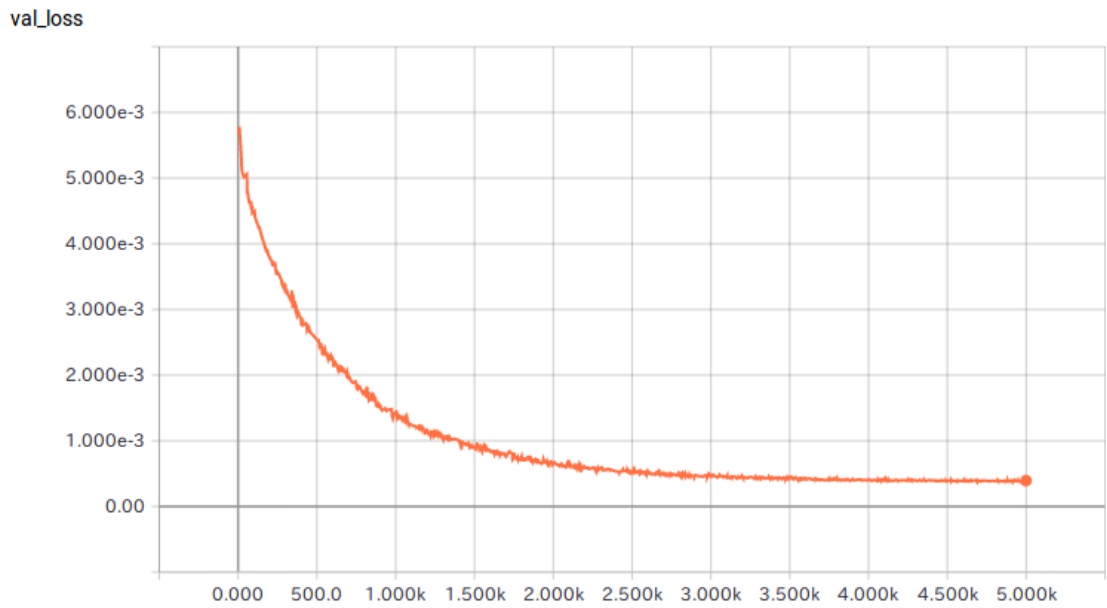


図 5.39 ch2 におけるテストデータの MSE(ネットワーク 3)



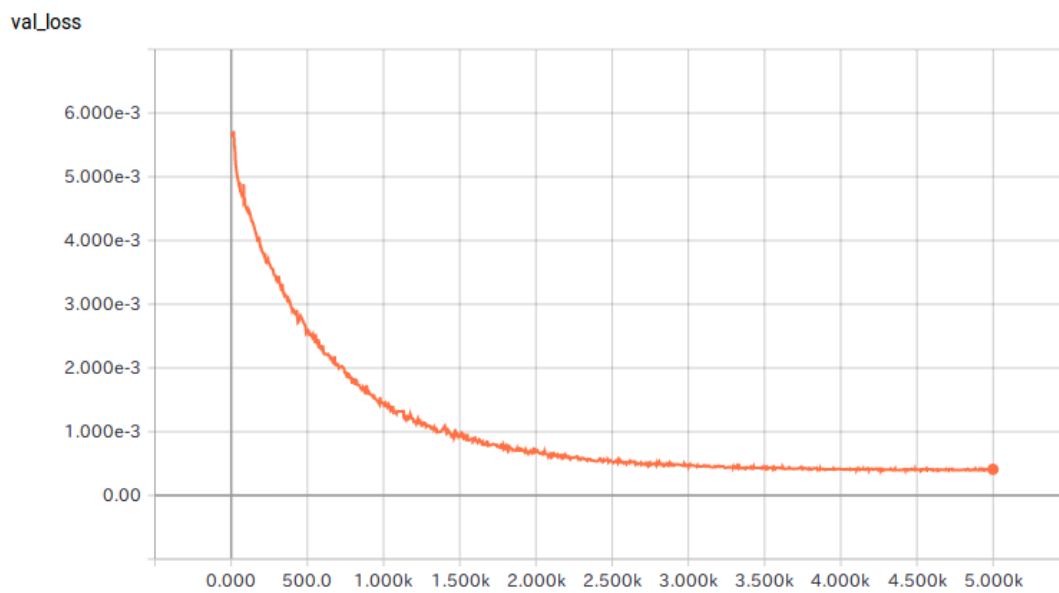


図 5.40 ch3 におけるテストデータの MSE(ネットワーク 3)

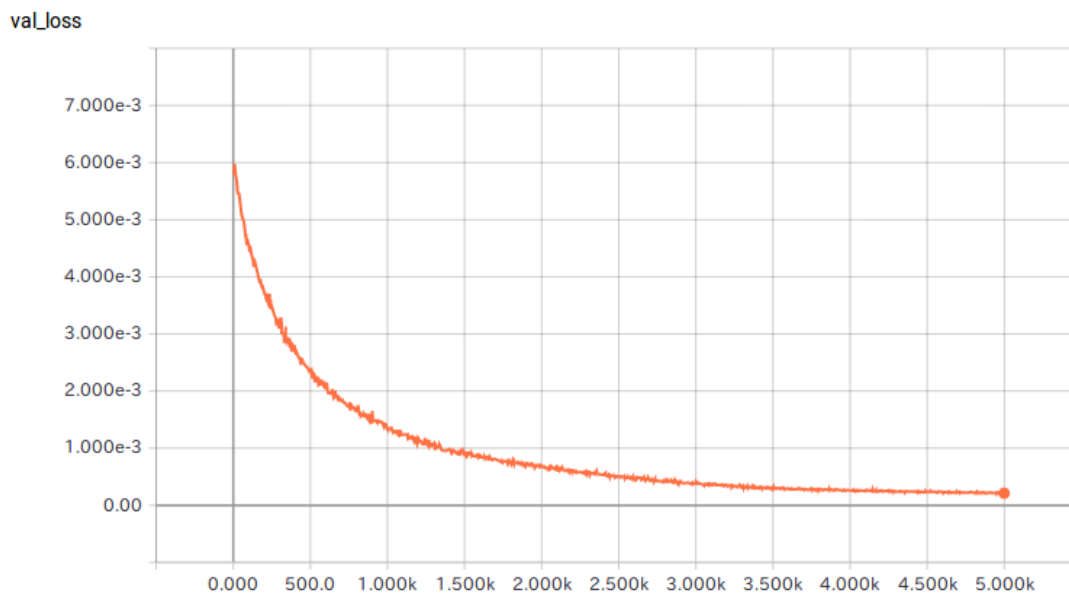


図 5.41 ch4 におけるテストデータの MSE(ネットワーク 3)

mean\_absolute\_percentage\_error

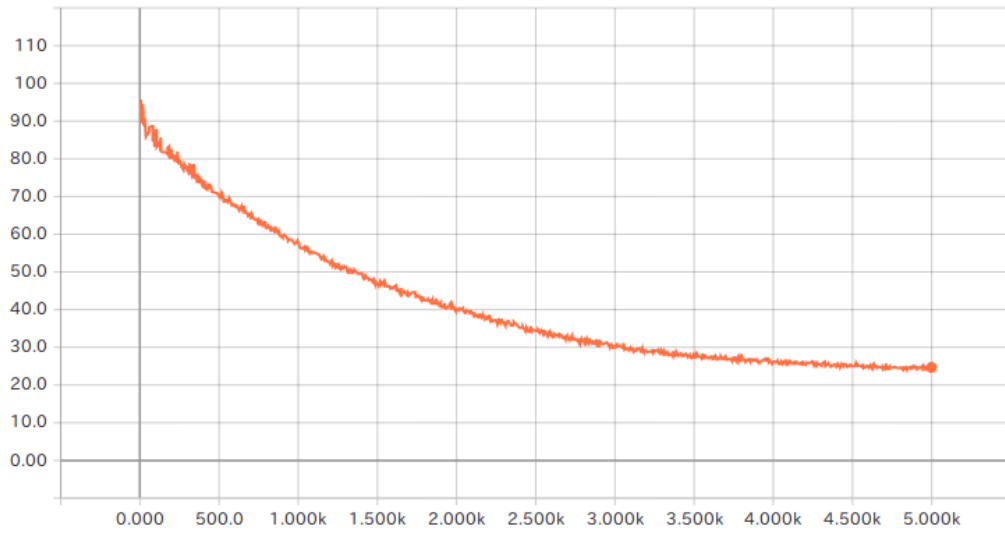


図 5.42 ch1 における学習データの MAPE(ネットワーク 3)

mean\_absolute\_percentage\_error

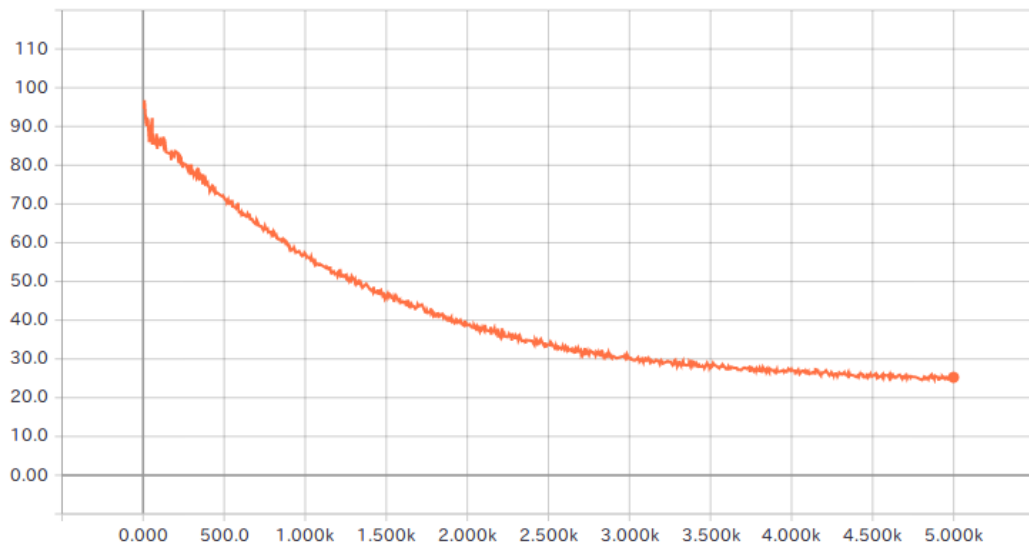


図 5.43 ch2 における学習データの MAPE(ネットワーク 3)

mean\_absolute\_percentage\_error

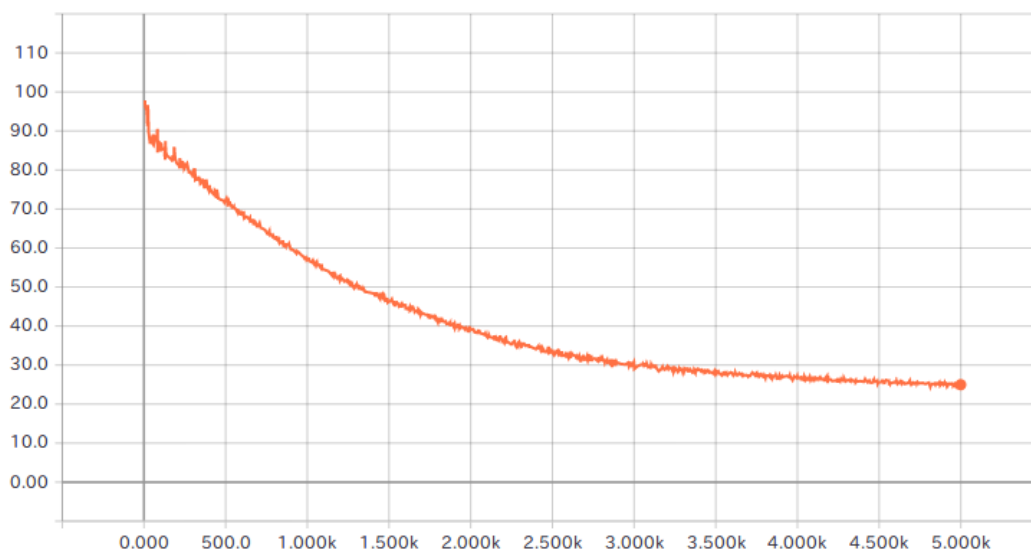


図 5.44 ch3 における学習データの MAPE(ネットワーク 3)

mean\_absolute\_percentage\_error

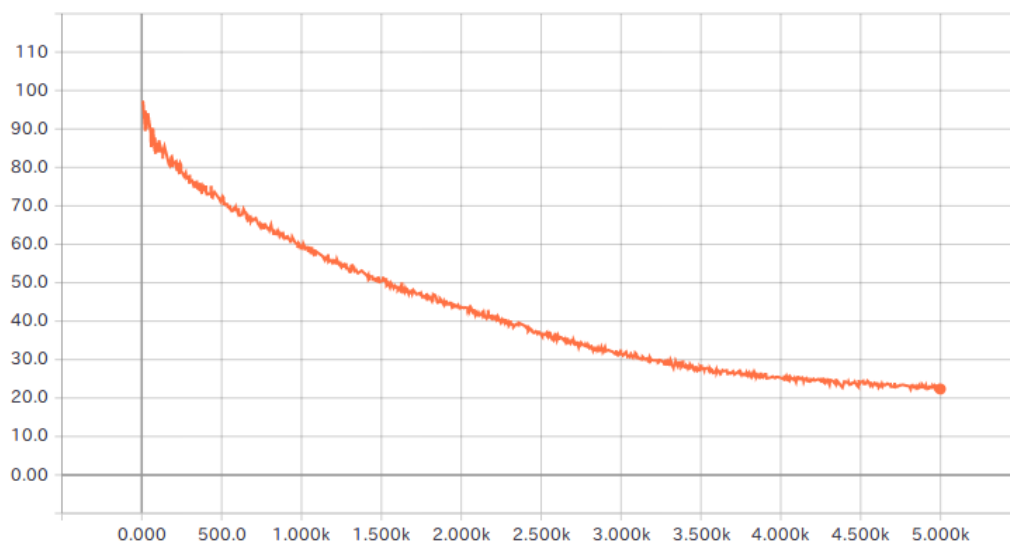


図 5.45 ch4 における学習データの MAPE(ネットワーク 3)

val\_mean\_absolute\_percentage\_error

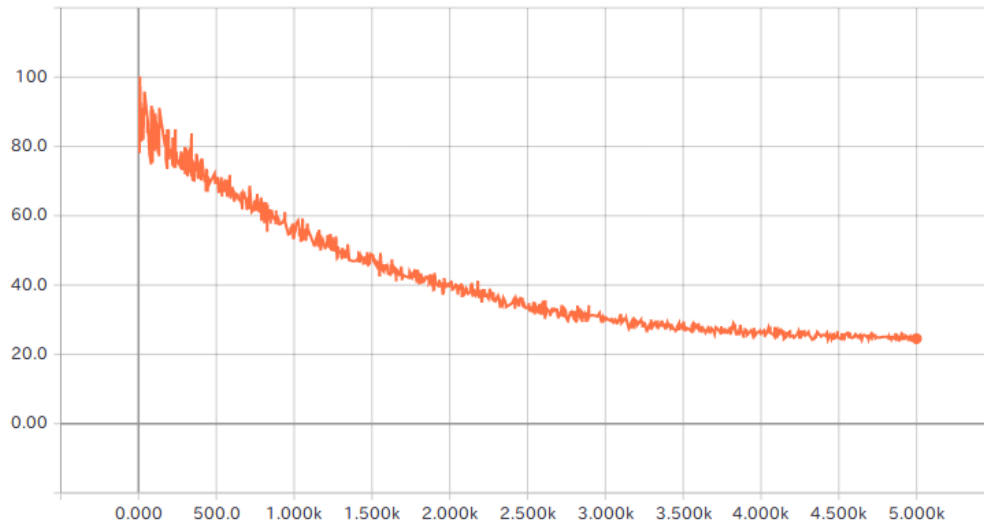


図 5.46 ch1 におけるテストデータの MAPE(ネットワーク 3)

val\_mean\_absolute\_percentage\_error

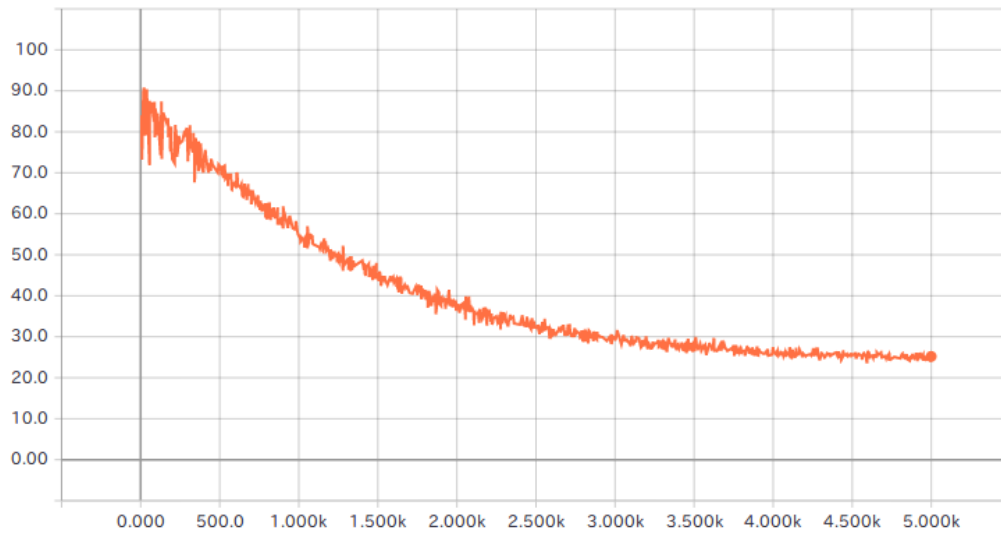


図 5.47 ch2 におけるテストデータの MAPE(ネットワーク 3)

val\_mean\_absolute\_percentage\_error

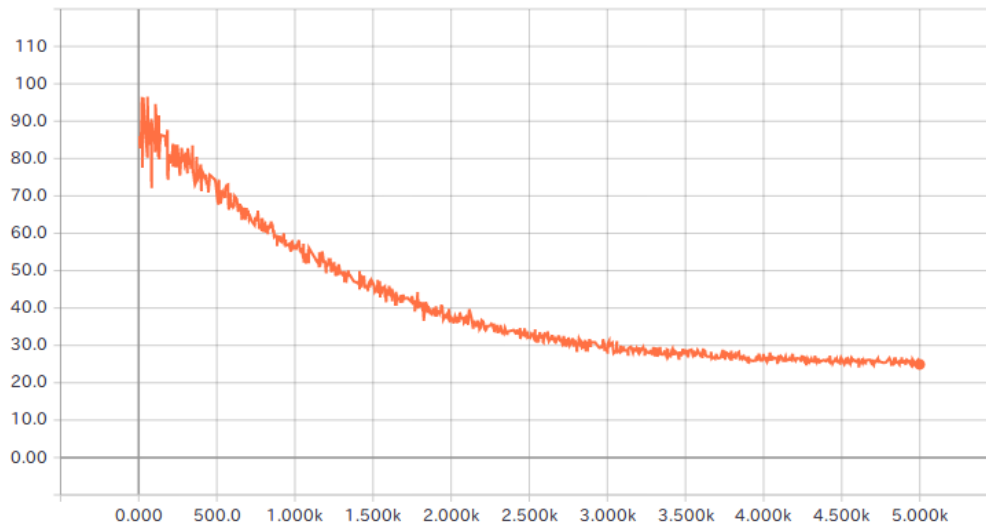


図 5.48 ch3 におけるテストデータの MAPE(ネットワーク 3)

val\_mean\_absolute\_percentage\_error

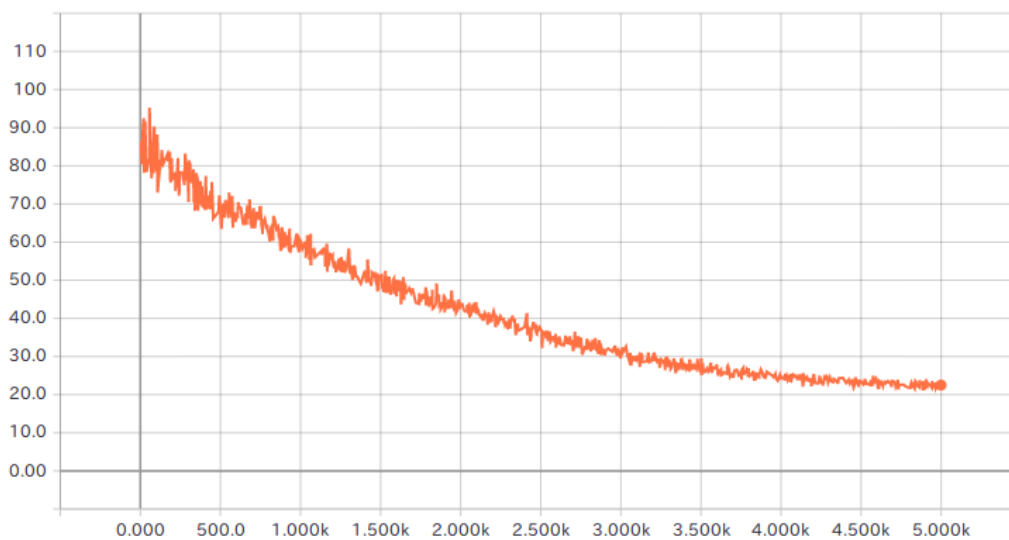


図 5.49 ch4 におけるテストデータの MAPE(ネットワーク 3)

図 5.34 から図 5.49 の結果から学習が収束していることが確認できる。MAPE の値を確認すると、すべてのチャンネルで 20-25 付近の精度で学習が収束していることが確認できる。よって、ch1 から ch4 の学習が成功しているといえる。

### 5.3 k-means 法によるクラスタリング

5.2 で学習を行った 3 種類のオートエンコーダで特徴表現の算出を行った。そこで得られた特徴表現に対して k-means 法によるクラスタリングを行った。3 種類のネット

ワークから得られた特徴表現に対して  $k=70$  を用いてクラスタリングを行った。

また、ネットワーク 2 およびネットワーク 3 で得られた特徴表現に対して XMeans を用いて適切な  $k$  の値の算出を行った。その結果、どちらも  $k=4$  であったため、ネットワーク 2、ネットワーク 3 から得られた特徴表現に対して  $k=4$  でのクラスタリングを行った。

これ以降、ネットワーク  $x$  から得られた特徴表現を  $k=y$  で  $k$ -means 法によるクラスタリングを行った結果を  $kmx-y$  と表記する。

なお、5.6 で使用した比較用音源を含む 205 音源分のデータに対してクラスタリングを行った。比較用音源の詳細は 5.6 に示す。

## 5.4 楽曲ごとの比較

それぞれのクラスタリング結果から 10 音源選択し、5 音源ずつプロットしたものを 2 枚ずつ示す。横軸はデータのインデックス番号を表している。インデックス番号は楽曲内での時系列を表しているが、音源の並び順は順不同となっている。縦軸は分類されたクラスタ番号、図下部の矢印は音源の変わり目を、図上部の T に続く数字は音源番号を表している。なお、10 音源は同じものを使用した。

$km1-70$  のものを図 5.50、図 5.51 に、 $km2-70$  のものを図 5.52、図 5.53 に、 $km3-70$  のものを図 5.54、図 5.55 に示す。また、 $km2-4$  のものを図 5.56、図 5.57 に、 $km3-4$  のものを図 5.58、図 5.59 に示す。

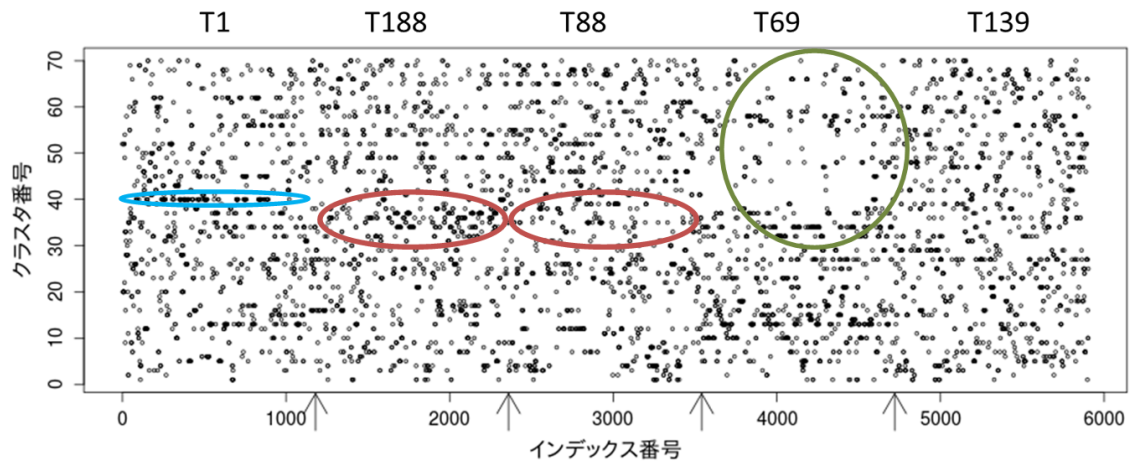


図 5.50 km1-70 における楽曲ごとの比較 1

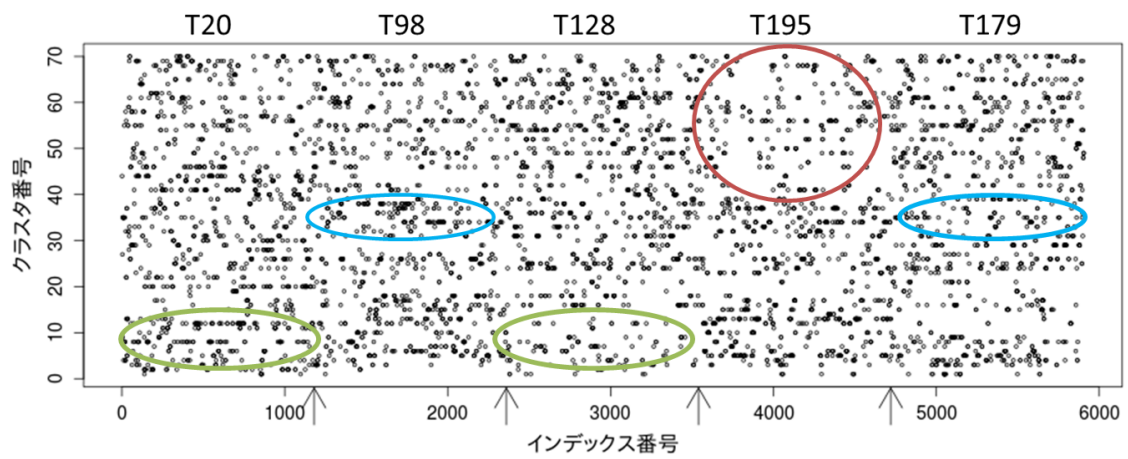


図 5.51 km1-70 における楽曲ごとの比較 2

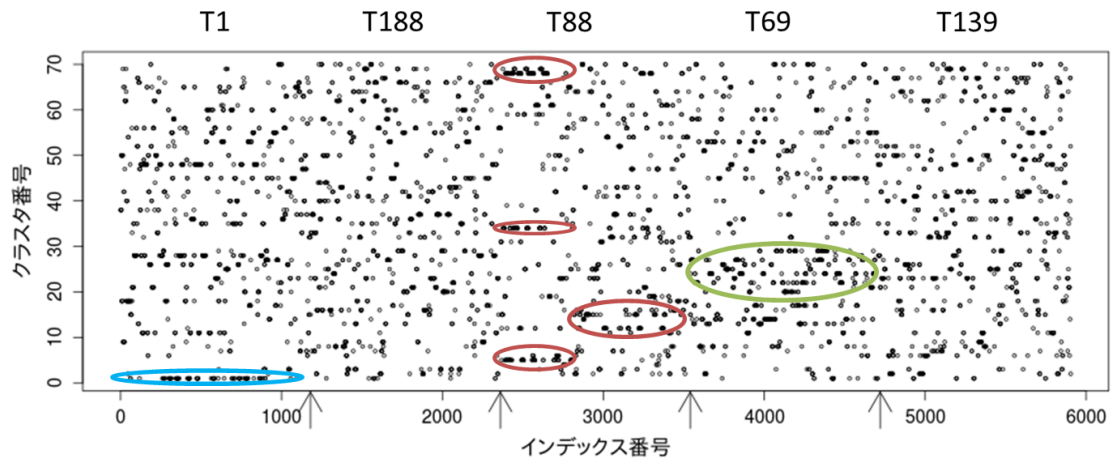


図 5.52 km2-70 における楽曲ごとの比較 1

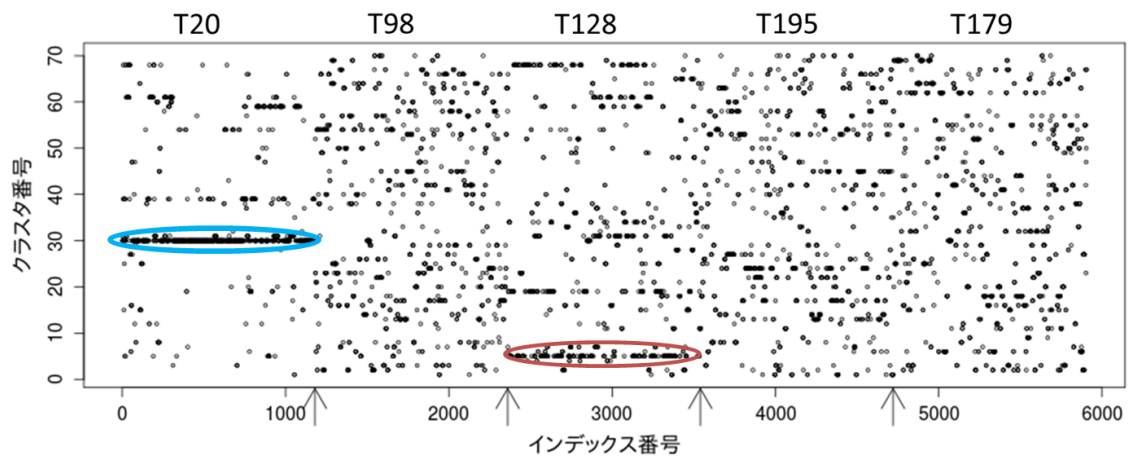


図 5.53 km2-70 における楽曲ごとの比較 2



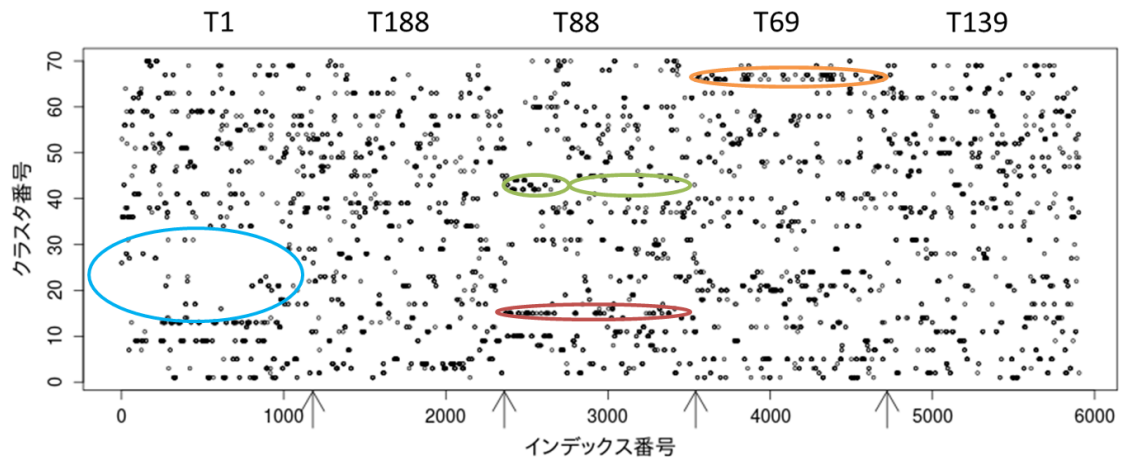


図 5.54 km3-70 における楽曲ごとの比較 1

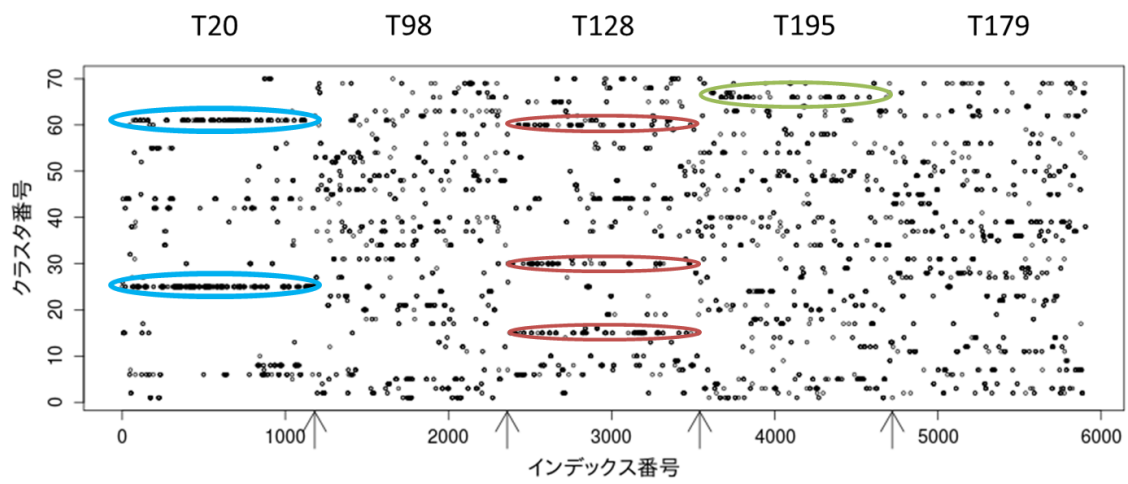


図 5.55 km3-70 における楽曲ごとの比較 2

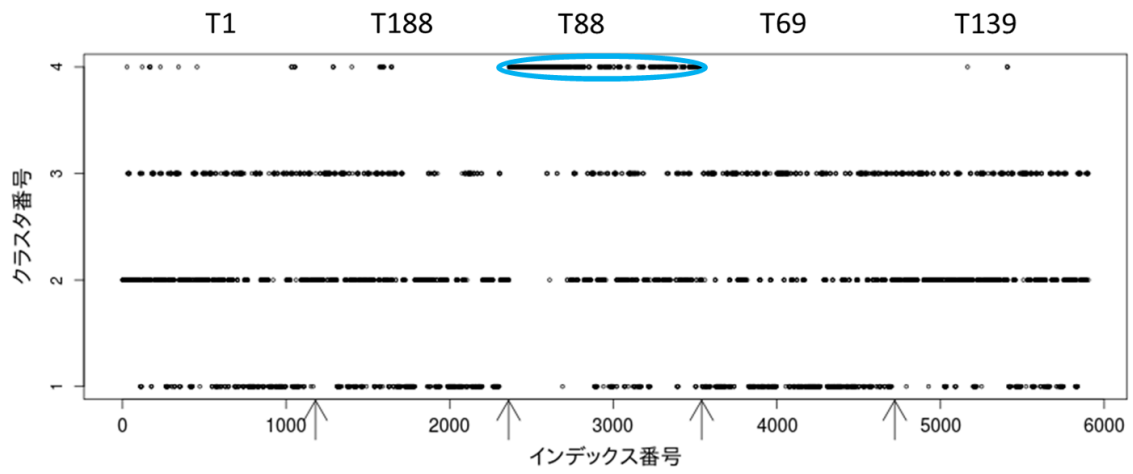


図 5.56 km2-4 における楽曲ごとの比較 1

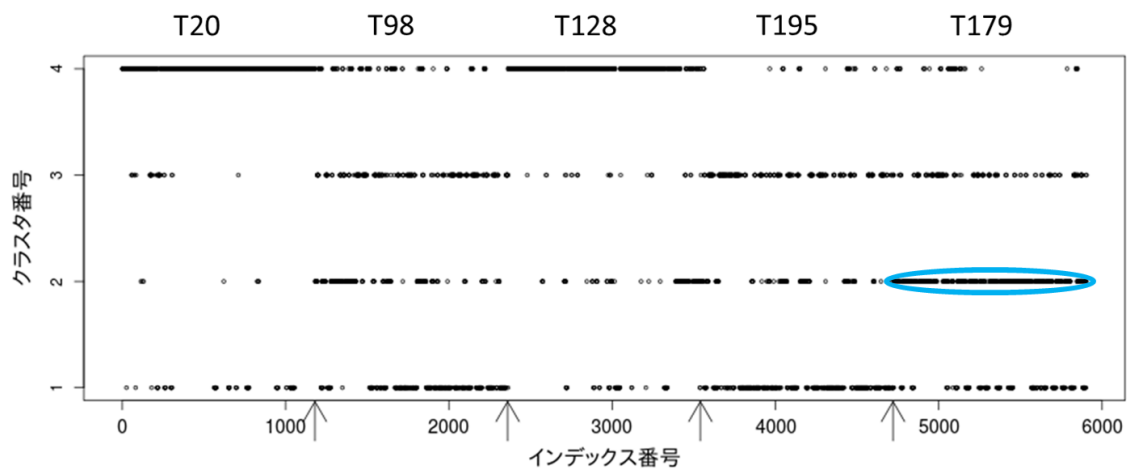


図 5.57 km2-4 における楽曲ごとの比較 2

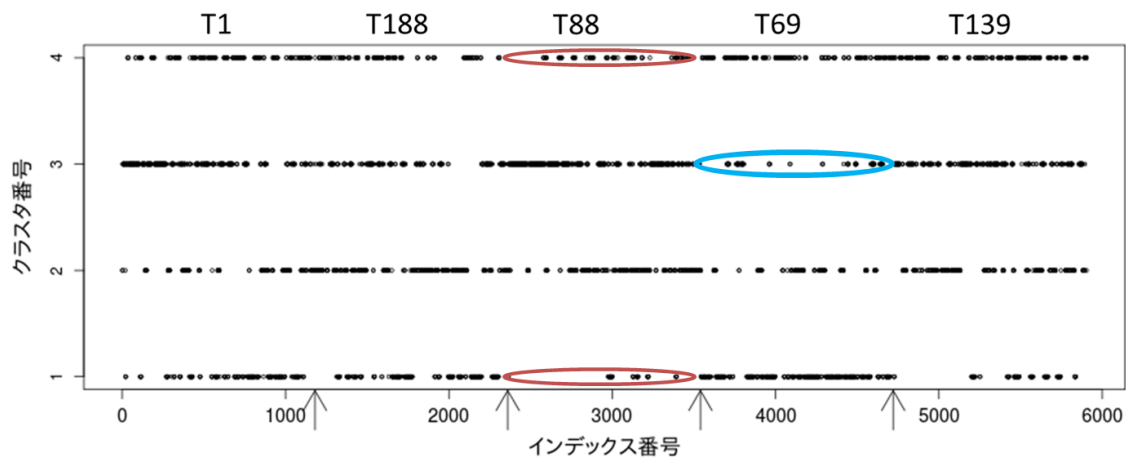


図 5.58 km3-4 における楽曲ごとの比較 1

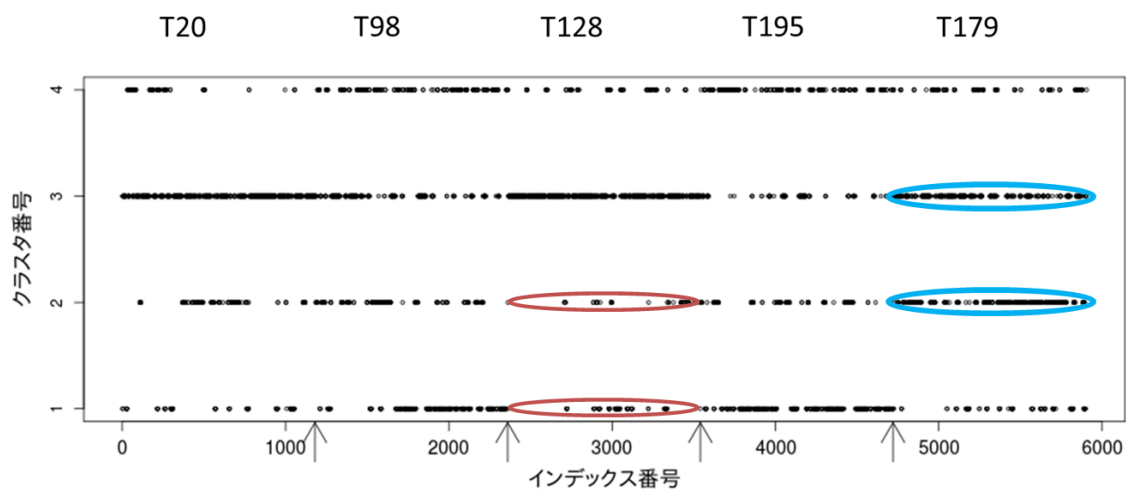


図 5.59 km3-4 における楽曲ごとの比較 2

上記の結果から、時間方向のクラスタの分類パターンが音源ごとに異なる特徴を持っているといえる。以下で結果ごとに細かく結果を見ていく。

km1-70 の結果について、図 5.50 の T1 ではクラスタ番号 40 付近にデータが密集している箇所が見られるが(青い円)、T69 では、クラスタ番号が 40 以上の箇所でデータが疎になっている(緑の円)。また、T188 と T88 を比較した際に、T188 の方がクラスタ番号 30-38 付近で密になっている(赤い円)。さらに、図 5.51 では、T98 のクラスタ番号 30-40 付近でデータが密になっている箇所が見られるが、T179 では同様の箇所では T98 に比べて疎になっている(青い円)。また、T20 と T128 を比較すると、T20 の方がクラスタ番号 1-15 付近でデータが密になっている(緑の円)。T195 では、クラスタ番号

40以上の箇所ではデータが疎に分布している(赤い円)。

km2-70の結果について、図 5.52 の T1 ではクラスタ番号 1 付近が密になっている箇所が見られる(青い円)。T88 では、インデックス番号 2800 付近(音源の前半)まではクラスタ番号 70、35、5 付近のデータが密になっているがそれ以降はクラスタ番号 15 付近が密になっている(赤い円)。T69 ではクラスタ番号 15-30 付近の箇所で密にデータが分布している(緑の円)。さらに、図 5.53 では、T20 のクラスタ番号 30 付近が非常に密になっていることが確認できる(青い円)。また、T128 ではクラスタ番号 8 付近が密になっていることが確認できる(赤い円)。

km3-70の結果について、図 5.54 の T1 ではクラスタ番号 15 から 30 付近の範囲でデータが疎になっている箇所が確認された(青い円)。T88 では km2-70 の場合と同様に緑の円の例のように、インデックス番号 2800 付近までとそれ以降では異なる分類パターンが観測されている。その一方で、クラスタ番号 18 付近では 1 音源通してデータが密に分布している(赤い円)。T69 ではクラスタ番号 65 付近でデータが密になっている等の特徴が観測された(オレンジの円)。図 5.55 では、T20 ではクラスタ番号 25 付近と、60 付近にデータが非常に密に分類されている(青い円)。T128 では、クラスタ番号 15、30、60 付近にデータが密に分類されていることが観測された(赤い円)。また、T195 では、クラスタ番号 65 付近でデータが密になっていることが観測された(緑の円)。

km2-4の結果について、図 5.56 において T88 のクラスタ 4 が密になっている傾向を持っている(青い円)。また、図 5.57 において T179 はクラスタ 2 が密になっている傾向を持っている(青い円)。このように、km2-70 と比較すると各音源の特徴は見づらいものの、異なる特徴を持っている音源が存在することが確認された。

km3-4の結果について、図 5.58 において、T69 では、クラスタ 3 が疎になっているという特徴を持っている(青い円)。一方で、T88 では、クラスタ 1 とクラスタ 4 が疎になっているという特徴を持つことが確認された。図 5.59 において T179 はクラスタ 2 とクラスタ 3 が密になっているという特徴を持っている(青い円)。また、T128 ではクラスタ 1 とクラスタ 2 が疎になっている(赤い円)。このように、km3-70 と比較すると各音源の特徴は見づらいものの、異なる特徴を持っている楽曲が確認された。

主に k=70 の場合、時間方向のクラスタの分類パターンが音源ごとに異なる特徴を持っていることが確認された。これらの特徴の違いは、音源による刺激または、それによって被験者の情動に変化が起きたことによって引き起こされた可能性が考えられる。

## 5.5 ジャンルごとの比較

それぞれのクラスタリング結果からメタル、J ポップ、演歌の音源を 4 音源ずつ選択し、各ジャンル 2 音源ずつプロットしたものを 2 枚ずつ示す。横軸はデータのインデックス番号を表している。インデックス番号は楽曲内での時系列を表しているが、音源の並び順は順不同となっている。縦軸は分類されたクラスタ番号、図上部の黒い矢印は音源の変わり目を、赤い矢印は音源及びジャンルの変わり目を表している。図上部の T に続く数字は音源番号を表している。なお、12 音源は同じものを使用した。

km1-70 のものを図 5.60、図 5.61 に、km2-70 のものを図 5.62、図 5.63 に、km3-70 のものを図 5.64、図 5.65 に示す。また、km2-4 のものを図 5.66、図 5.67 に、km3-4 のものを図 5.68、図 5.69 に示す。

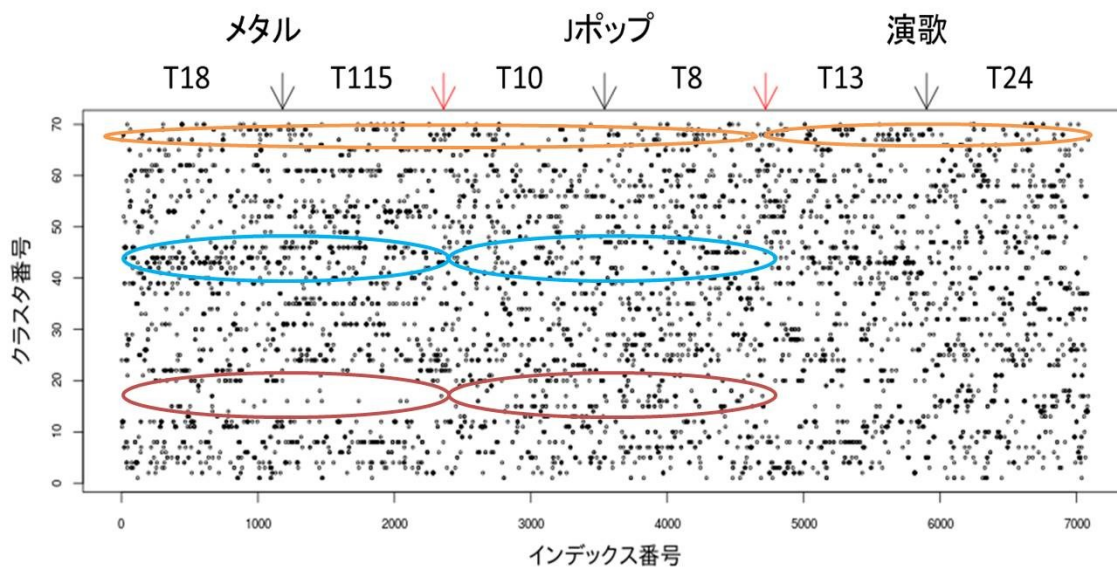


図 5.60 km1-70 におけるジャンルごとの比較 1

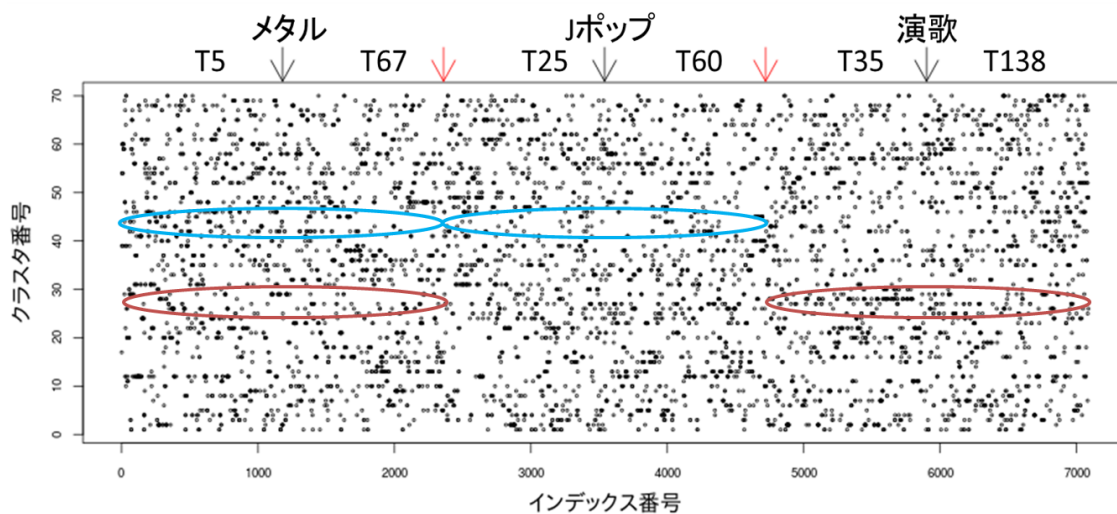


図 5.61 km1-70 におけるジャンルごとの比較 2

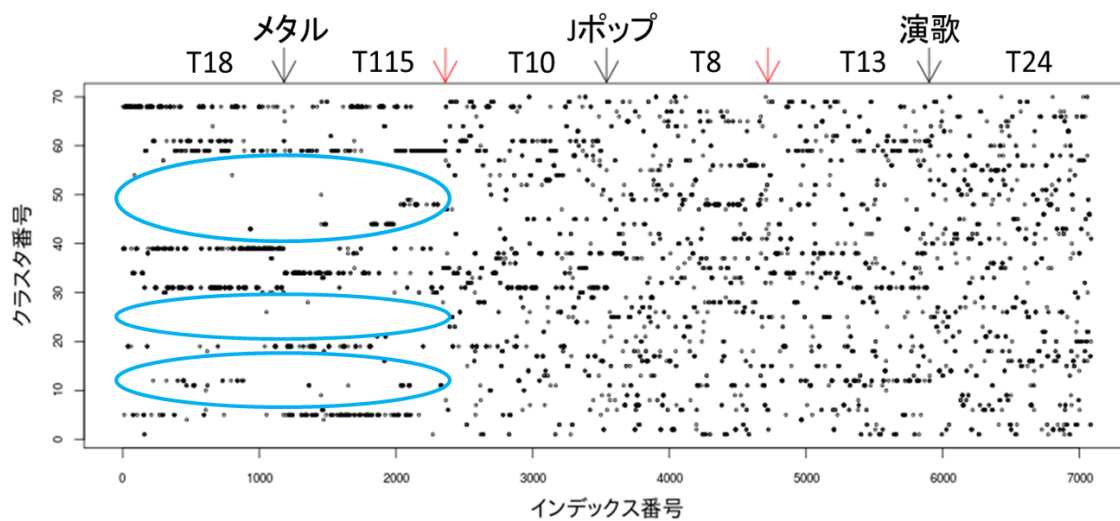


図 5.62 km2-70 におけるジャンルごとの比較 1

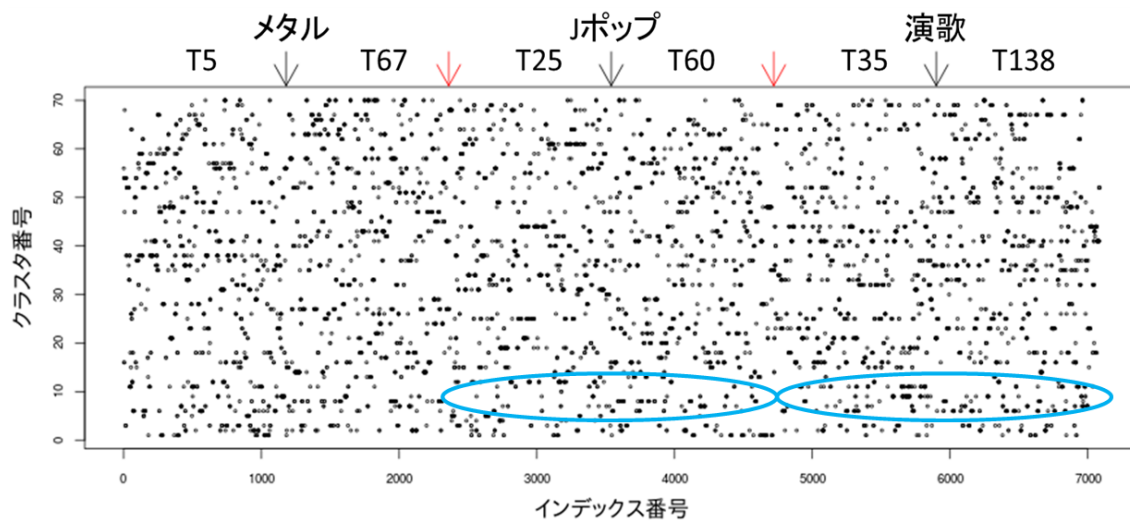


図 5.63 km2-70 におけるジャンルごとの比較 2

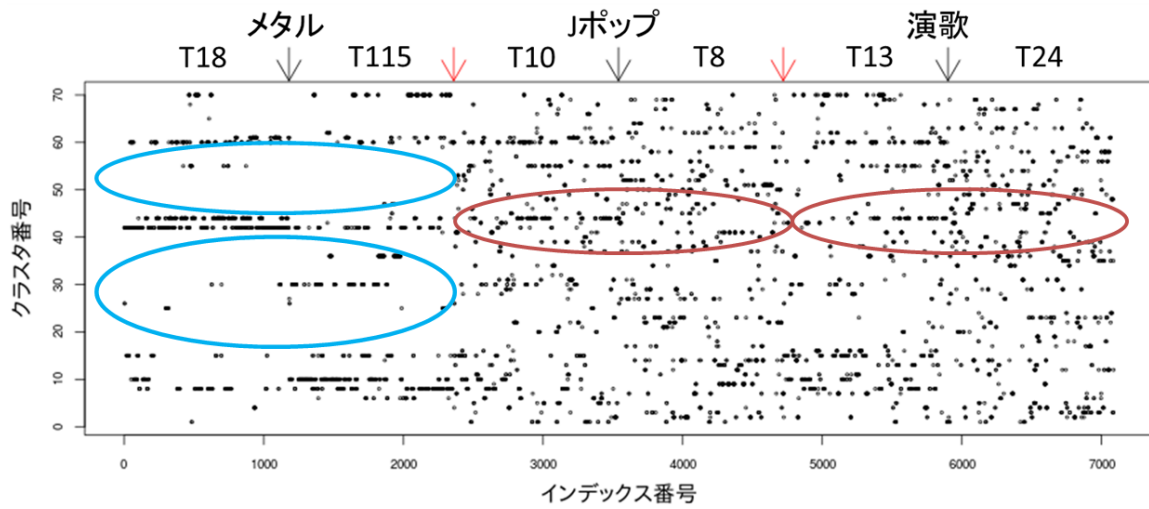


図 5.64 km3-70 におけるジャンルごとの比較 1

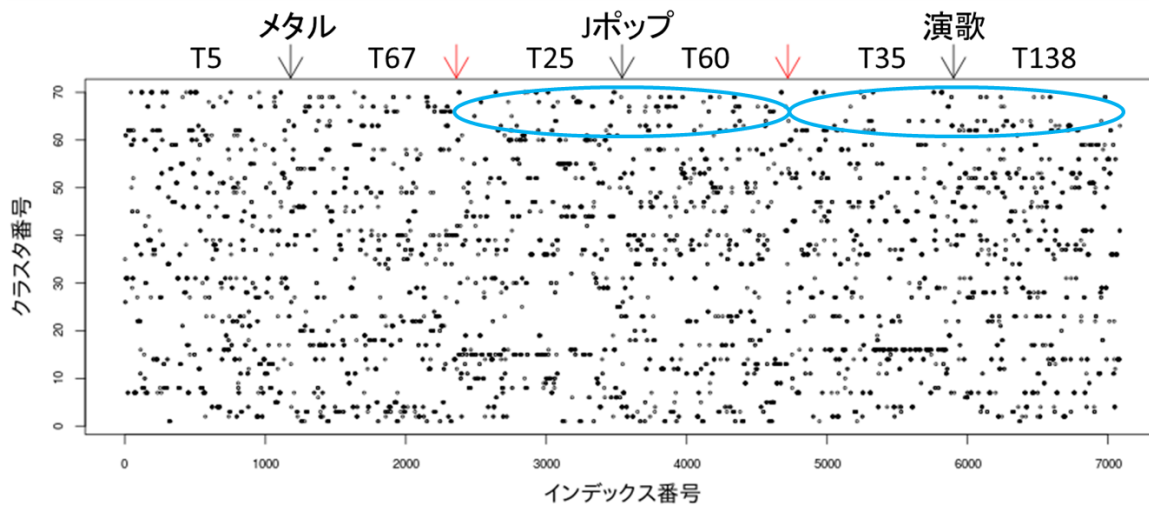


図 5.65 km3-70 におけるジャンルごとの比較 2



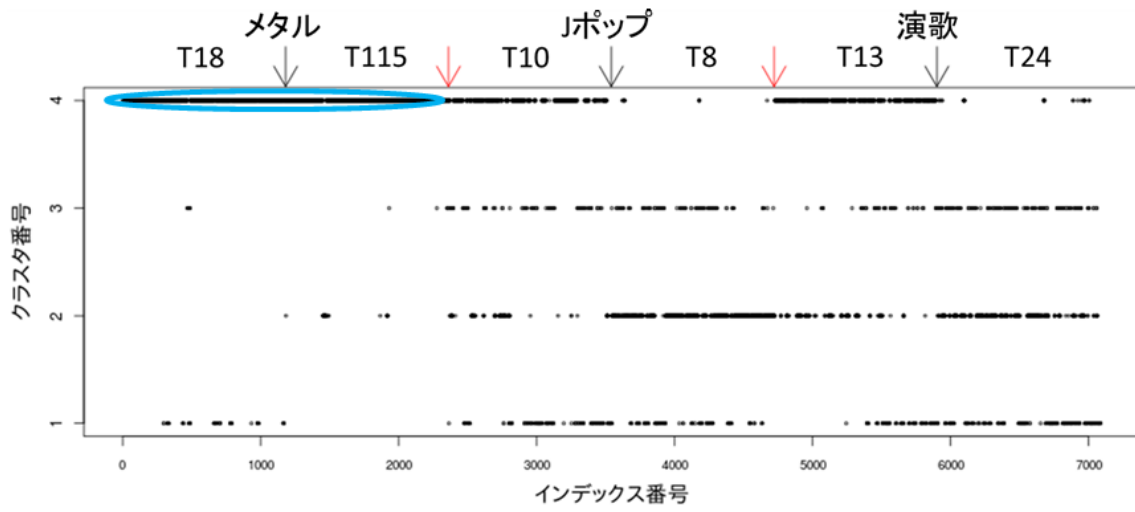


図 5.66 km2-4 におけるジャンルごとの比較 1

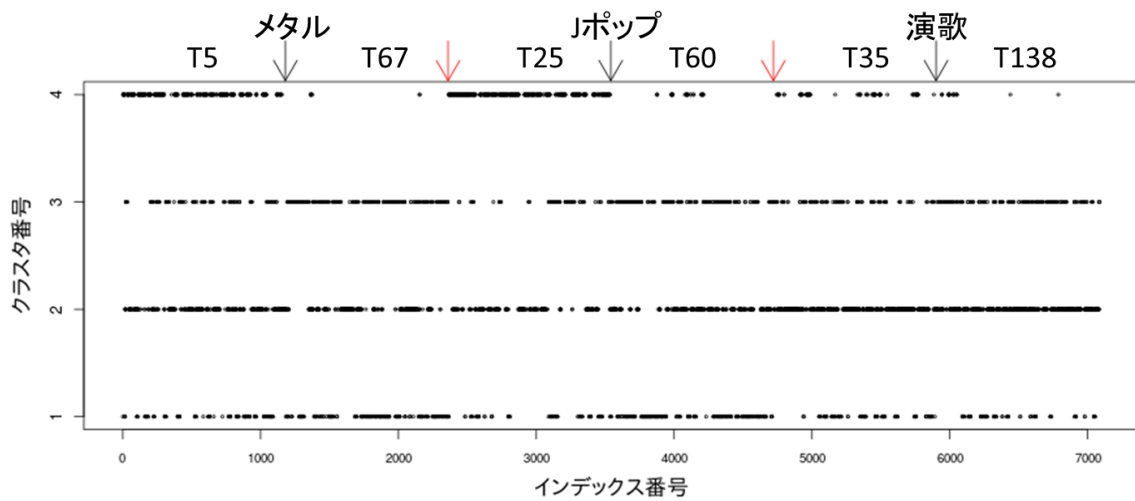


図 5.67 km2-4 におけるジャンルごとの比較 2

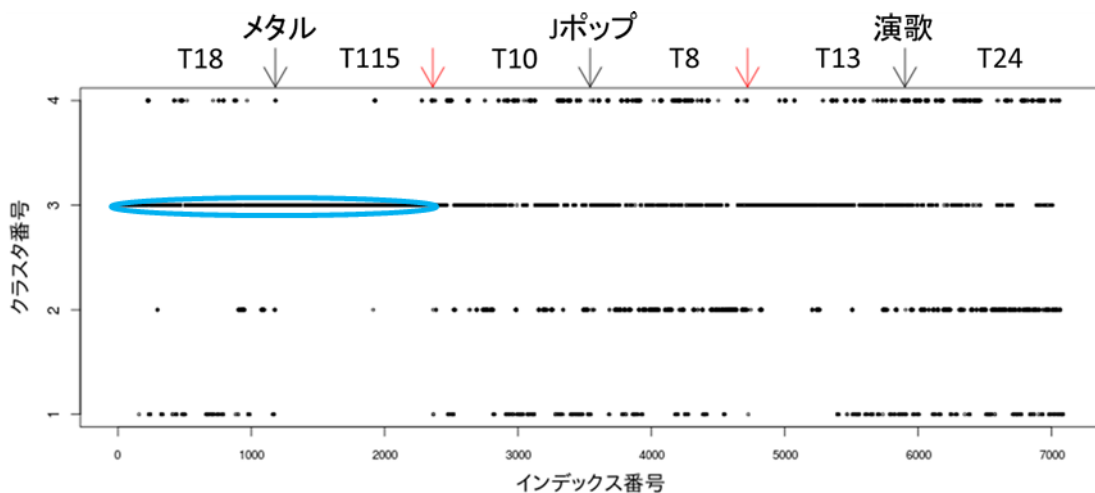


図 5.68 km3-4 におけるジャンルごとの比較 1

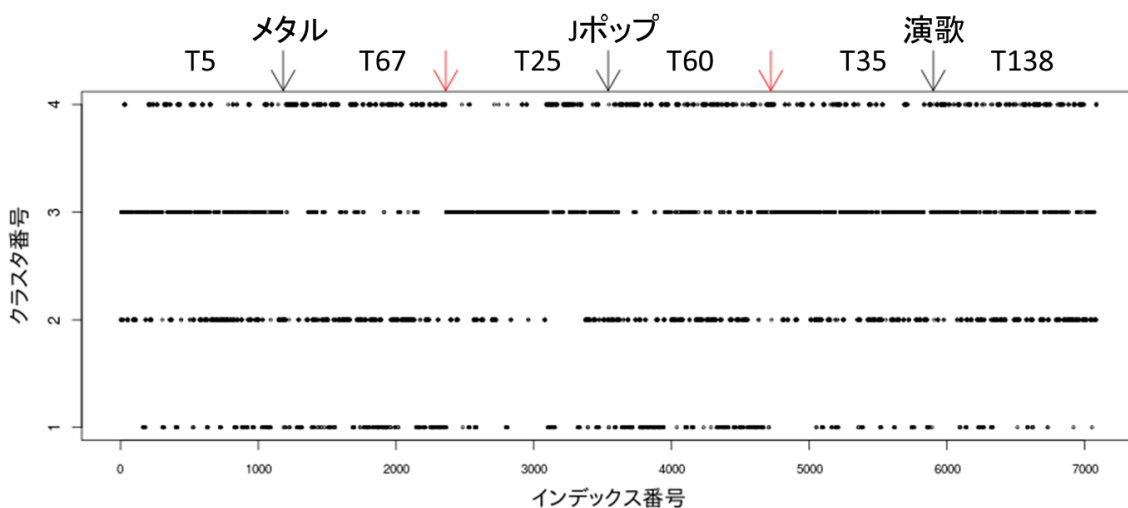


図 5.69 km3-4 におけるジャンルごとの比較 2

クラスタリング結果の時間方向の分類パターンがジャンル毎に異なる特徴を持っていることが確認できる。

km1-70 では図 5.60 の結果について、メタルと J ポップを比較した場合、メタルの方がクラスタ番号 40-45 付近でデータが密に(青い円)、15-20 付近で疎になっている(赤い円)。また、演歌ではクラスタ番号 70 付近で比較的データが密になっているなどの特徴がある。図 5.61 の結果について、図 5.60 よりもジャンルごとの傾向は見づらいものの、メタルと J ポップを比較した際にクラスタ 40-45 付近がメタルの方が密にデータが分布していることが確認できる。また、メタルと演歌を比較するとクラスタ番号

25-30 付近において、演歌の方がメタルよりもデータが密に分布している傾向が見られた。

km2-70 について、図 5.62 では、青い円で囲った箇所が大きく疎になっており、メタルの 2 音源がほかのジャンルと明らかに異なる特徴を持っていることがわかる。また、図 5.63 の J ポップと演歌を比較するとクラスタ番号 5-12 付近で演歌の方が、データが密に分布している(青い円)。また、図 5.62 と図 5.63 のメタルを比較すると、異なる特徴を持っていることがわかる。

km3-70 について、図 5.64 においてもメタルの 2 音源がほかのジャンルと異なる特徴を持っていることが確認できた。また、J ポップと演歌を比較するとクラスタ番号 38-50 付近の区間で J ポップの方が演歌よりもわずかにデータが密に分布していることが確認できる(赤い円)。図 5.65 において、J ポップと演歌を比較した際にクラスタ番号 60-70 付近で J ポップの方が、データがわずかに密に分類されていることが確認できる(青い円)。また、km2-70 の場合と同様、図 5.64 と図 5.65 のメタルを比較すると、異なる特徴を持っていることがわかる。

k=4 として k-means 法によるクラスタリングを行った図 5.66 から図 5.69 に関しては、図 5.66 のメタルの 2 音源では、青い円で囲った部分にデータが密に分布していること、また、それ以外のクラスタが非常に疎になっていることが確認できる。図 5.68 のメタルの 2 音源では、青い円で囲ったクラスタが他のジャンルの該当部分よりもデータが密になっており、その他のクラスタが疎になっていることが確認された。それ以外は、図からはジャンルごとの特徴が判別しづらいという結果になった。

上述のように、ジャンルごとにクラスタリング結果の時間方向の分類パターンに特徴がある場合が見られた。その一方で、同一ジャンルでも共通の特徴が見つけない場合も見られた。これは、同じジャンルでも曲調が異なる楽曲が存在することや、聴取時の被験者の情動や体調などのコンディションの違いなどに起因していると考えられる。

文献[5]においても、ジャンルによって異なる基礎律動パターンが観測されること、同じジャンルの楽曲でも異なる基礎律動パターンが現れる場合があることが報告されているが、異なる手法を用いて検証を行った場合でも報告されている内容と同様の結果が観測された。

また、メタルの T18、T115 の 2 音源の分類傾向が km2 と km3 ではデータが少数のクラスタに密に分布しているのに対し、km1 では km2、km3 と比較すると全クラスタに渡ってデータが分散しているという結果が観測された。km1 では 1-30[Hz]を使用しており、km2、km3 では 1-45[Hz]を使用して学習を行っている。そのため、ネットワークを学習する際に使用した基礎律動データの周波数が異なることに起因して分類傾

向の違いが表れている可能性がある。これは、差分の 15[Hz]に音源の特徴に関する情報が含まれている可能性や、ノイズが含まれているためこのような結果になった可能性がある。また、オートエンコーダのノード数などの層構造の違いに起因する可能性も考えられる。しかし、どの要因が影響してこのような結果が得られたのかについての検討はできていない。ネットワーク 2 とネットワーク 3 の構造を比較すると、ネットワーク 3 においては LSTM を利用しており、入力データを時系列データとして扱っている点異なる。そのため、結果に大きな差が得られると予測していたが、実際に km2 と km3 を比較した際に、メタルの T18、T115 の 2 音源の分類傾向を比較すると、少数のクラスタにデータが密に分類される傾向が共通しており、大きな分類傾向の違いは確認できなかった。その原因として分類の際に時間方向のデータの量が重要でない可能性や、ノイズのようなランダムな成分が多く時間方向の変動を正確に学習できていない可能性などが考えられる。

## 5.6 同一音源を複数回聴取した場合の比較

先行研究[7]では基礎律動を用いて楽曲の聴取順序が推定可能であると報告されているが、聴取時期が異なる同一の音源を聴取した際に、同一のパターンが見られるかどうかの検証を行った。

検証に際して、初回聴取から約 1 年後に 200 音源の中から 5 音源を選択し、再度聴取を行った。その後オートエンコーダを用いた特徴表現の算出、k-means 法によるクラスタリングを行った。

インデックス番号は楽曲内での時系列を表している。縦軸は分類されたクラスタ番号、図下部の矢印は音源の変わり目を表している。T1 と T7 の 2 曲分の比較を図 5.70 から図 5.79 に示す。

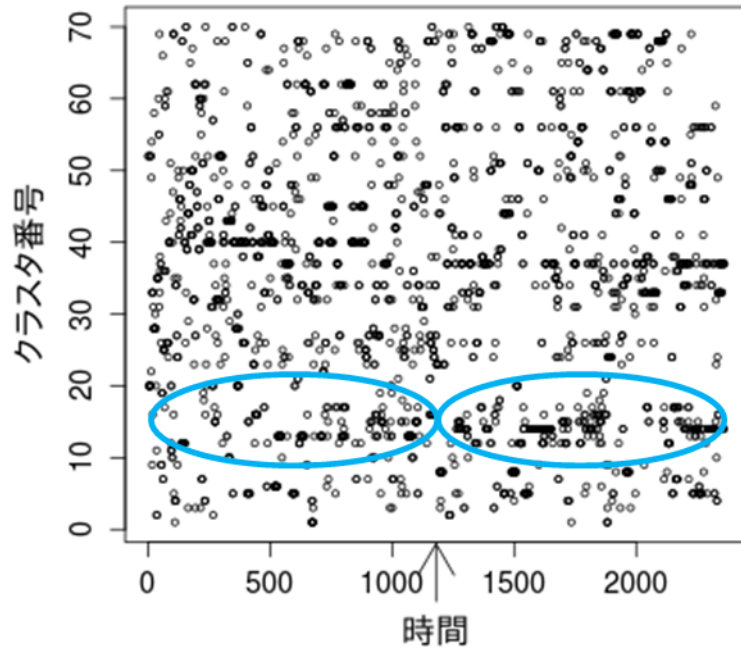


図 5.70 km1-70 において T1 を複数回聴取した場合の比較

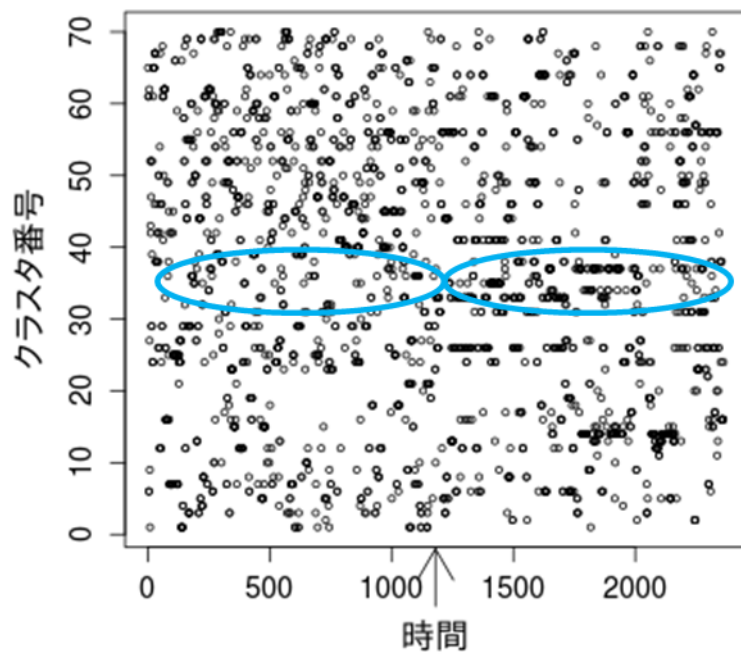


図 5.71 km1-70 において T7 を複数回聴取した場合の比較

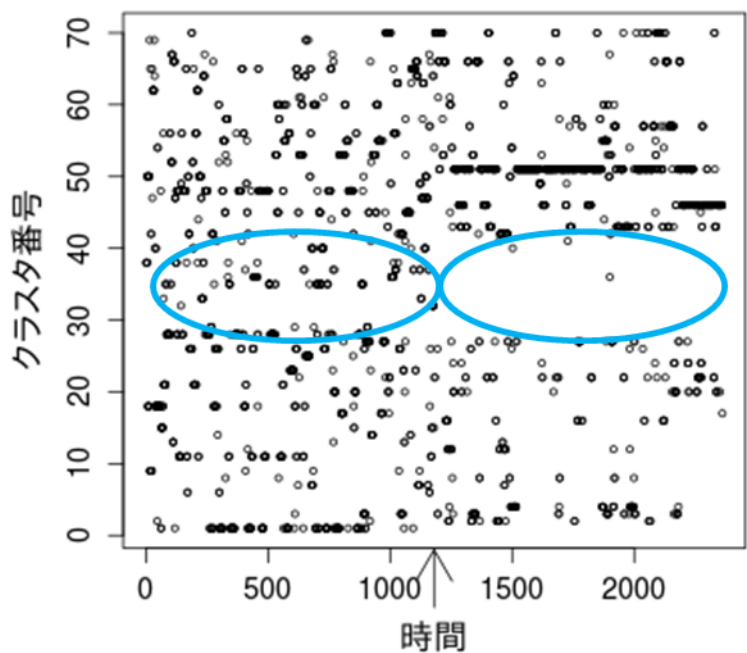


図 5.72 km2-70 において T1 を複数回聴取した場合の比較

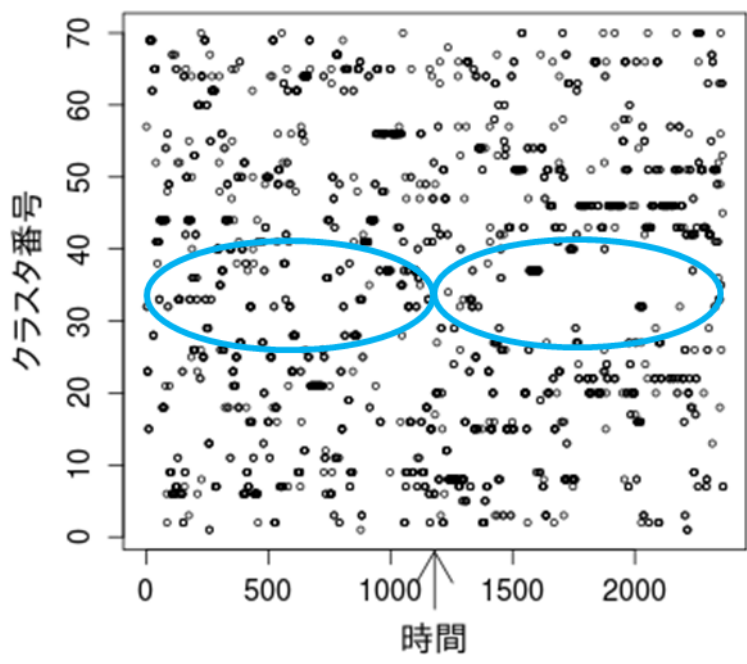


図 5.73 km2-70 において T7 を複数回聴取した場合の比較

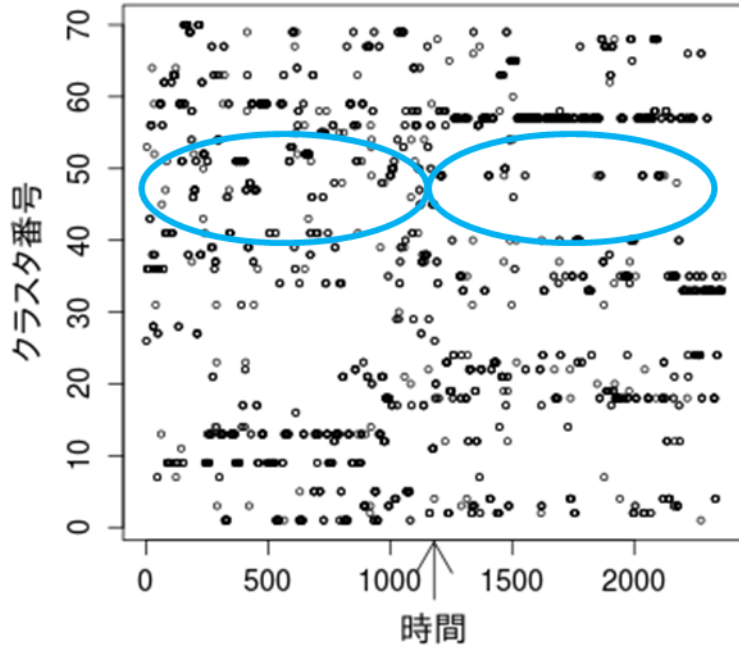


図 5.74 km3-70 において T1 を複数回聴取した場合の比較

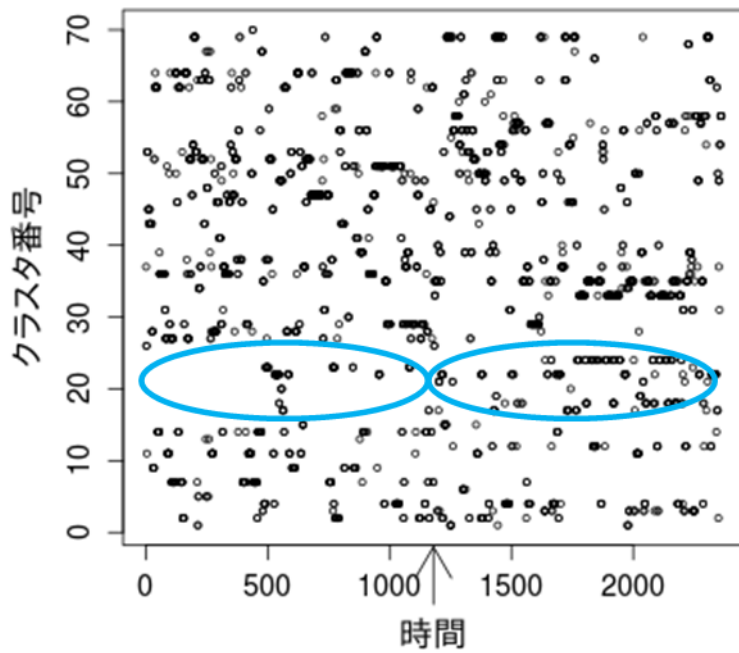


図 5.75 km3-70 において T7 を複数回聴取した場合の比較

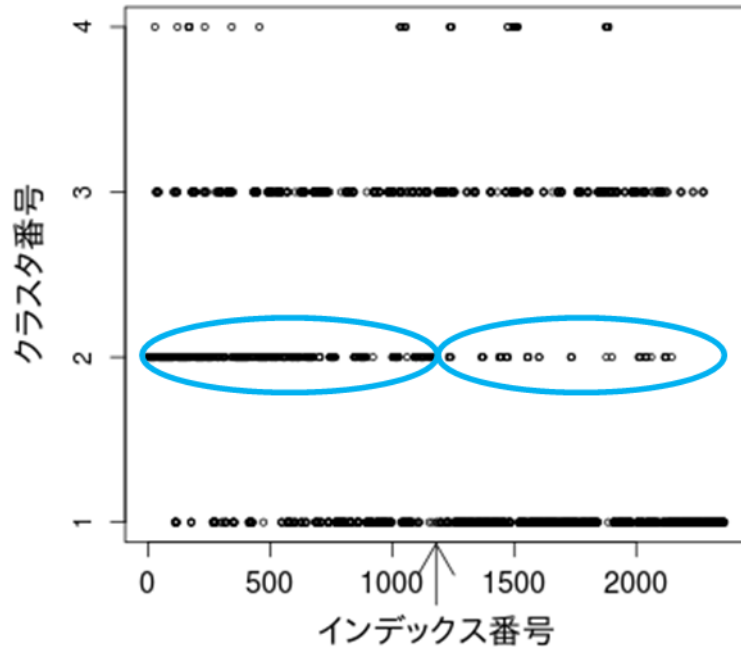


図 5.76 km2-4 において T1 を複数回聴取した場合の比較

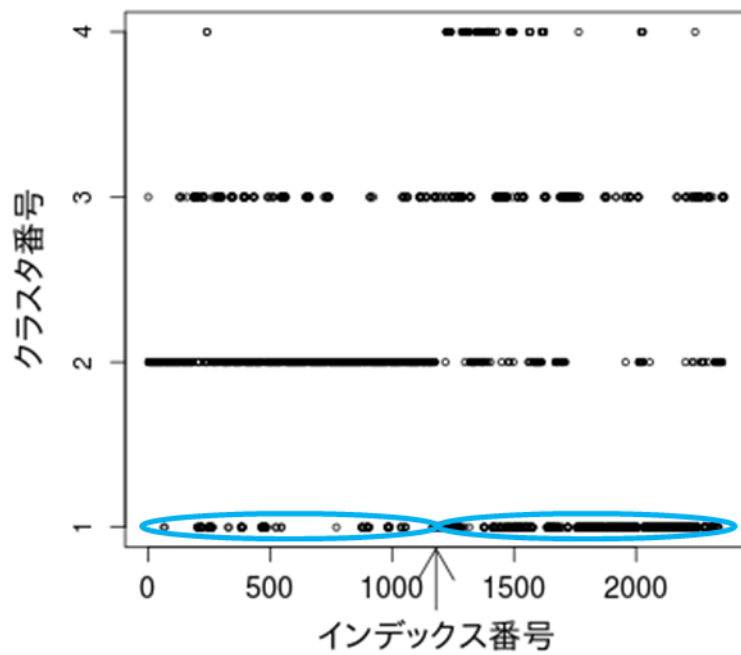


図 5.77 km2-4 において T7 を複数回聴取した場合の比較



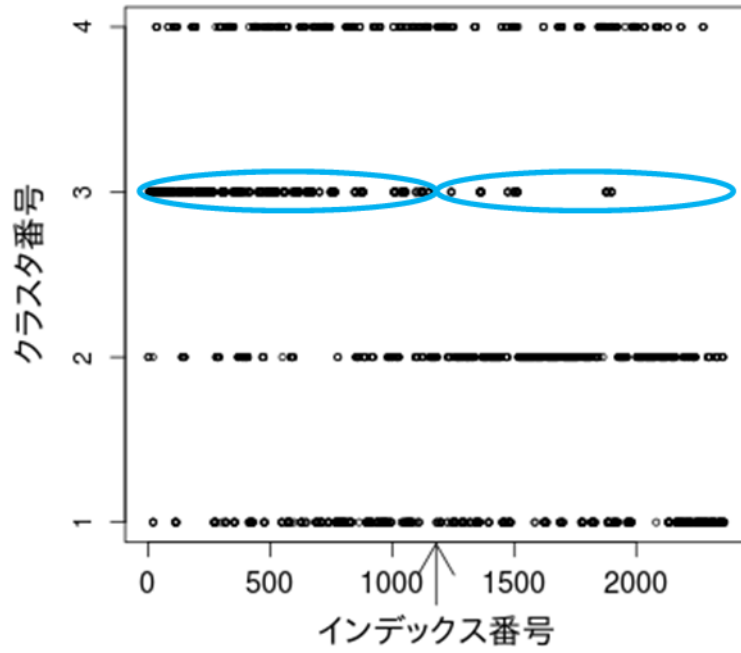


図 5.78 km3-4 において T1 を複数回聴取した場合の比較

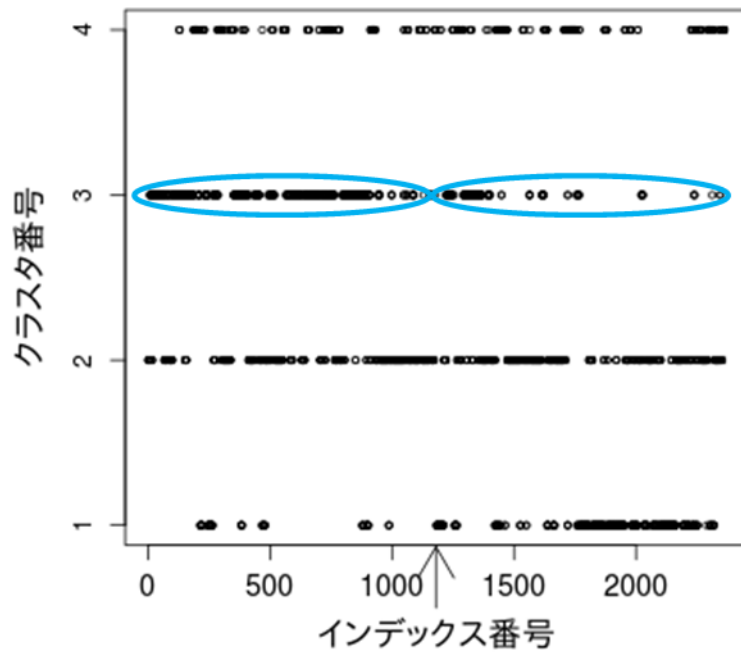


図 5.79 km3-4 において T7 を複数回聴取した場合の比較

図 5.70 から図 5.79 で比較した 2 音源に関して、それぞれの図で青い円で囲った例のように、同一音源を聴取した場合でも時間方向のクラスタリング結果の分類パターンが異なることが確認された。今回の検証では、複数回聴取を行ったデータにおいて共通した特徴は見つけれなかった。文献[7]の中で、楽曲聴取中の基礎律動には順序効果がある可能性が報告されているが、本検証の結果からも同様の結果が観測された。しかし、さらに細かく調査を行った場合や、k-means 法やオートエンコーダのパラメータを適切に調節した場合に聴取時期が異なる同一の音源の間でも共通の特徴が見いだせる可能性があるため、今後も検討していく必要があるといえる。

## 5.7 クラスタリング結果の定量的な評価手法の検討

5.3 で行ったクラスタリングの結果に対して、5.4 から 5.6 までは実際にプロットし評価するという定性的かつ主観的な評価手法で評価を行ってきた。この章では定量的に評価する手法の検討を行った。

### 5.7.1 手法 1

5.3 で行ったクラスタリングの結果に対し 1 音源当たりの各クラスに分類されたデータ数を集計し、それを特徴ベクトルとして再度クラスタリングを行うことで、5.3 のクラスタリング結果の分類傾向が近いものをグループ化した。2 回目のクラスタリングにはウォード法を用いた階層的クラスタリングを行った。

2 回目のクラスタリングを行った後、クラスタリング結果のデンドログラムの近い音源の一部を被験者に再度聴取させ、楽曲の類似性があるかどうかの確認を行った。さらに、それぞれの印象評定項目ごとに実験時に回答してもらったアンケートの回答結果との確認を行った。

以下に 2 回目のクラスタリング結果を表したデンドログラムを示す。

図中の点線の中の音源に対して被験者による再聴取、アンケートの回答結果との確認を行った。

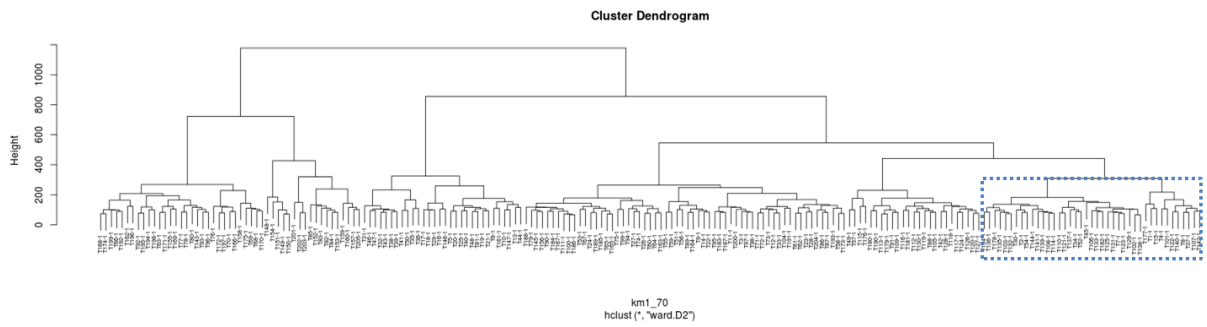


図 5.80 km1-70 における 2 回目のクラスタリング結果

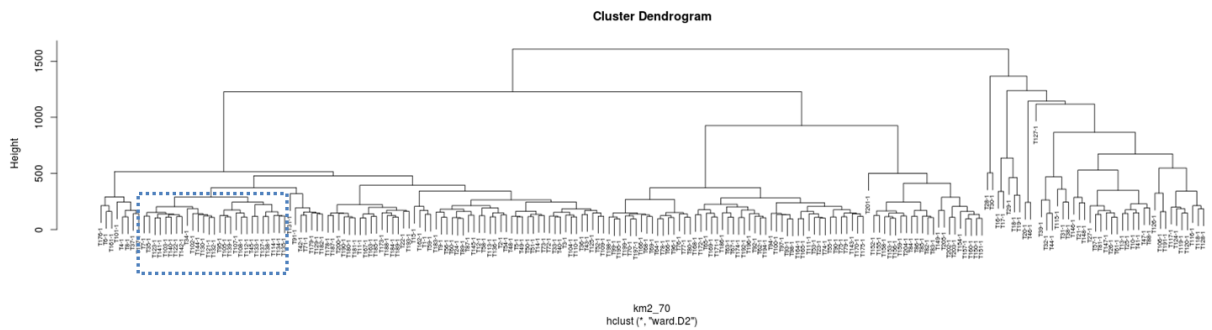


図 5.81 km2-70 における 2 回目のクラスタリング結果

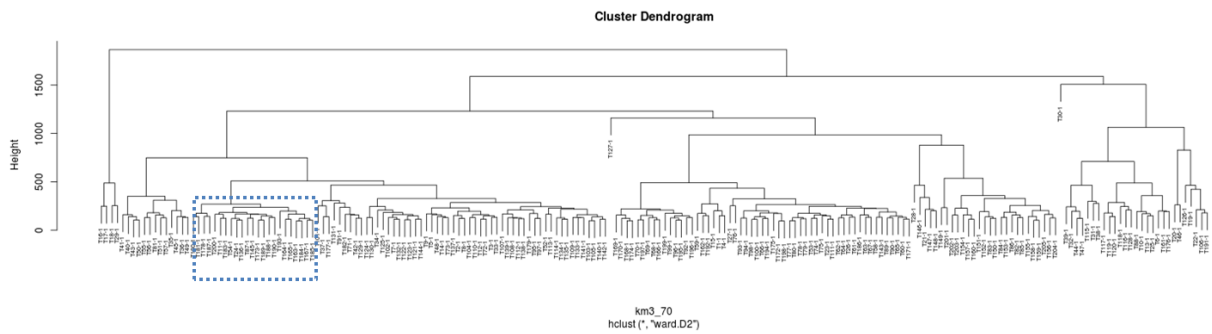


図 5.82 km3-70 における 2 回目のクラスタリング結果

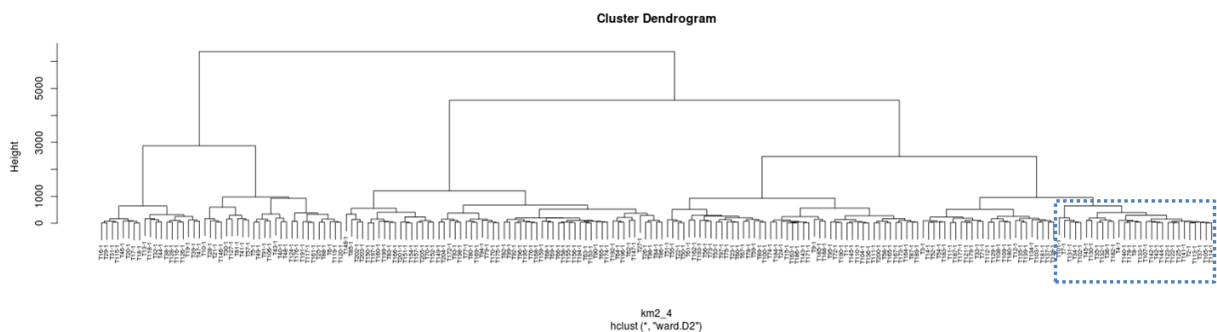


図 5.83 km2-4 における 2 回目のクラスタリング結果

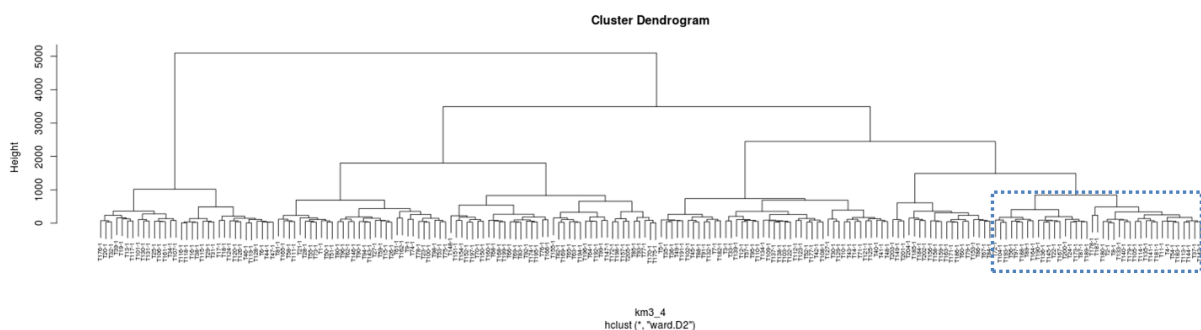


図 5.84 km3-4 における 2 回目のクラスタリング結果

被験者にクラスタリング結果における点線で囲った部分を被験者に聴取させたところ、すべてのクラスタリング結果において楽曲のジャンルにおける偏りは確認できなかった。また、テンポが速い曲や遅い曲、明るい曲や暗い曲が確認でき、分類結果に一貫性は見いだせなかった。

次に、点線で囲まれた部分をアンケートの回答結果と対応させて調査を行ったところ、結果から共通した特徴は見いだせなかった。

k-means 法によるクラスタリング結果を再度クラスタリングする際に、1 曲のデータすべてを集計してしまったため、楽曲内での時間情報が消えてしまい、時間方向の変動を考慮できずに、被験者の主観評価とは異なる分類結果になってしまった可能性が考えられる。

## 5.7.2 手法 2

手法 2 では、グラフ分析を用いてクラスタリング結果の評価を試みた。グラフ分析とは、グラフの中で、つながれているノード間の関係性に着目して解析を行う解析手法である。

グラフ分析を行うための前処理として、クラスタリング結果に関するグラフの作成を行った。k-means 法のクラスタリング結果を音源 6 秒分のデータごとに 1 音源当たり 10 個に分割し、それぞれの各クラスターに分類されたデータの数の集計を行い、特徴ベクトルの作成を行う。その後、ベクトル間のユークリッド距離が近いもの 5%をつなぐ形でグラフの作成を行った。

そうして作成されたグラフに対してデータ研磨を行った。データ研磨とは、ある条件を満たすノード間にはリンクを張り、一方で満たさないノード間のリンクは切るという処理を行うアルゴリズムである。データ研磨を行うことで与えられたグラフを、そのグループ構造を壊さずにより明確な構造にできるとされている[14]。データ研磨には、Github で公開されているライブラリ[15]を使用した。

データ研磨後のグラフから、大きさが 3 であるクリークを抽出した。抽出したものの中から一部を再度被験者に聴取させることで評価を行った。

実際に被験者に聴取させたものを表 5.4 から表 5.8 に示す。表の楽曲はTx-yという形式であらわされているが、音源番号xのy番目を表す。なお解析対象は 6 秒分ごとだが、6 秒で音源の評価を行うことは難しいと予想されたため、被験者への聴取は 1 音源すべてで行った。

表 5.4 km1-70 で被験者に聴取させた音源

1	T146-4,T32-3,T47-7
2	T56-3,T50-9,T51-1
3	T133-4,T30-6,T112-10
4	T76-8,T70-8,T168-7
5	T45-3,T72-3,T169-1

表 5.5 km2-70 で被験者に聴取させた音源

1	T32-5,T115-3,T176-2
2	T148-8,T154-3,T157-9
3	T104-2,T169-6,T199-7
4	T142-1,T36-6,T130-8
5	T121-10,T35-9,T102-10

表 5.6 km3-70 で被験者に聴取させた音源

1	T7-5,T79-9,T145-7
2	T93-4,T83-6,T163-2
3	T179-8,T75-4,T80-2
4	T93-2,T145-10,T174-5
5	T96-1,T73-8,T166-4

表 5.7 km2-4 で被験者に聴取させた音源

1	T195-5,T97-8,T188-2
2	T37-10,T7-9,T186-8
3	T173-9,T59-9,T164-8
4	T57-6,T22-9,T91-4
5	T43-5,T38-4,T176-9

表 5.8 km3-4 で被験者に聴取させた音源

1	T2-5,T95-7,T184-3
2	T2-6,T69-9,T148-4
3	T3-7,T98-1,T180-3
4	T7-4,T20-6,T176-8
5	T8-8,T167-6,T170-4

上記の楽曲を被験者に聴取させたところ、3音源すべてに共通する特徴が見つけれない場合が多いという回答が得られた。このことから、この手法の有効性は確認できなかった。手法1と異なり、手法2では解析対象が6秒のみであり、解析対象区画と被験者が聴取した範囲に差があったため正確ではない結果になってしまった可能性が考えられる。

## 5.8 まとめと考察

5章では、聴取楽曲及び聴取ジャンルと基礎律動との間に対応関係があるかどうかの検証を行った。5.2、5.3でオートエンコーダの学習、特徴表現の算出、k-means法によるクラスタリングを行った。5.4では、音源ごとに基礎律動データのクラスタリング結果の時間方向における分類パターンの比較を行い、音源ごとに異なる基礎律動パターンが観測された。5.5では、ジャンルごとに基礎律動データのクラスタリング結果の時間方向における分類パターンの比較を行い、ジャンルごとに分類パターンに特徴が見られる場合があることが確認された。5.6では、時間を空けて同一の音源を聴取した場合に共通のパターンが見いだせるかどうかの検証を行った。その結果、今回検証を行った範囲では共通のパターンを見出すことはできなかった。5.7ではクラスタリング結果に対してより定量的な評価を試みたが、被験者の主観による評価や、アンケート結果に合致する結果は得られなかった。5.4、5.5、5.6の結果から楽曲と基礎律動パターンとの間には対応関係が存在し、その関係は1対1の対応関係ではなく、1対多の関係がある可能性が考えられる。

5.4および5.5から、基礎律動と聴取した楽曲および、ジャンルとの間に対応関係がある可能性が示唆される。また、5.6から同一楽曲を聴取した場合の基礎律動パターンは複数ある可能性が考えられる。基礎律動パターンが複数現れる要因として、実験時における被験者の体調などの身体的な要因や、音源聴取時の被験者情動や実験時のストレスなどの心理的な要因や、音源を聴取する際に楽器を意識するのか、ボーカルを意識するのか、リズムを意識するのかなどの音楽の聴き方などが要因として考えられる。しかし、今回の検証ではどの要因の影響で基礎律動パターンが複数現れているのか、原因の断定はできなかった。

## 第6章 結論

### 6.1 まとめ

本研究では、音楽聴取中の基礎律動を用いて、楽曲聴取者が楽曲に対して持った印象を推定することを目標として、基礎律動データに対して教師あり学習および教師なし学習を用いて分類を行った。

教師あり学習を用いてアンケートの回答結果との対応の検証を行った。SVM を使用した検証では、前処理の方法にかかわらず外挿の分類精度がチャンスレベル程度にとどまった。ディープラーニングを用いて行った検証ではテストデータの分類精度が 0.35 から 0.5 付近を境に過学習が起きてしまい、うまく学習することができなかった。

教師なし学習では、オートエンコーダを用いて 3 種類のネットワークを用意して中間表現の算出を行い、k-means 法によるクラスタリングを行った。その結果を定量的に評価するために、1 音源当たりの各クラスに分類されたデータ数の集計を行い、これを特徴ベクトルとして再度ワード法による階層型クラスタリングを行うことにより、k-means 法の結果における分類傾向が近いもののグループ化を行った。その結果、同一グループ内の楽曲でも様々な曲調の楽曲が確認され、分類結果に一貫性が見いだせなかった。そこで、実験時に回答されたアンケートとの比較を行ったが、共通した特徴は見いだせなかった。グラフ分析を用いて行った解析でも、被験者の主観による評価に合致するような結果は得られなかった。その一方で、k-means 法によるクラスタリング結果と聴取楽曲との対応に関する検討を行った結果、k-means 法のクラスタリング結果が聴取楽曲ごとに異なる時間方向の分類パターンを持っていることが確認できた。また、k-means 法のクラスタリング結果とジャンルとの対応の検討を行った際に、クラスタリング結果の時間方向の分類パターンがジャンルごと異なるものが存在することが確認された。これらの結果から、基礎律動が聴取楽曲及び聴取ジャンルと対応関係を持っている可能性が示唆された。一方で、同一楽曲を、時間を空けて聴取した場合の比較を行った検証では、同一音源を聴取した場合でも異なるクラスタリング結果の分類パターンが確認された。このことから、基礎律動と楽曲との対応関係は必ずしも一致しない、つまり、基礎律動と楽曲との対応関係は 1 対多である可能性が示唆された。

これらの結果から、基礎律動と聴取楽曲及びジャンルとの間に対応関係がある可能性があること、基礎律動と楽曲との対応関係は 1 対多である可能性が示唆された。



## 6.2 今後の課題

### 6.2.1 ハイパーパラメータの調整

4.3 で学習を行った際に、すべての検証で過学習が起きてしまっていた。そのため、過学習が起きないようにパラメータ調整を行い、十分なエポック数の学習を行うことで精度の改善が見られる可能性がある。また、5.2 で学習を行った際に、学習自体は収束したが、その際に得られた特徴表現が分類を行うのに最適でなかった可能性が考えられる。そのため、ハイパーパラメータを適切に調整することでより被験者の印象や感じ方に近い分類が行えるようになる可能性がある。

### 6.2.2 クラスタリング結果に対する評価手法の検討

今回は k-means 法のクラスタリング結果に対して再度クラスタリングを行うことでクラスタリング結果の評価を試みた。しかし、この手法では結果に一貫性が見いだせなかった。そのため、より効果的なクラスタリング結果の評価手法の検討が必要である。さらに、今回提案した手法によってジャンルや楽曲ごとに異なるクラスタリング結果の分類パターンが確認された。しかし、これらの分類パターンがどのような要因で分類されているのかなどの検討ができていないため、今後検討を行っていく必要がある。

### 6.2.3 アンケートの改善

90 秒の音源を聴取させているのに対し、アンケートは 1 つのみをとっているため、楽曲の変動に対して、アンケートの窓幅が広すぎる可能性が考えられる。楽曲の印象は A メロ、B メロ、サビなどで変化する可能性があるため、アンケートの時間方向の分解能を向上させる必要がある。

### 6.2.4 被験者数の不足

今回の実験では被験者が男性 1 名であり、結果がその被験者特有のものである可能性が考えられる。そのため、被験者数を増やし、他の被験者でも同様の結果が得られるか検討する必要がある。

## 図一覧

図 3.1 muse の電極位置 .....	5
図 3.2 実験の流れ .....	10
図 3.3 アンケート集計 .....	12
図 4.1 検証 5 の学習結果(accuracy) .....	18
図 4.2 検証 5 の学習結果(loss).....	18
図 4.3 検証 6 の学習結果(accuracy) .....	20
図 4.4 検証 6 の学習結果(loss).....	20
図 4.5 検証 7 の学習結果(accuracy) .....	22
図 4.6 検証 7 の学習結果(loss).....	22
図 4.7 検証 8 の学習結果(accuracy) .....	24
図 4.8 検証 8 の学習結果(loss) .....	24
図 5.1 オートエンコーダ学習の概要図.....	27
図 5.2 ch1 における学習データの MSE(ネットワーク 1).....	28
図 5.3 ch2 における学習データの MSE(ネットワーク 1).....	28
図 5.4 ch3 における学習データの MSE(ネットワーク 1).....	29
図 5.5 ch4 における学習データの MSE(ネットワーク 1).....	29
図 5.6 ch1 におけるテストデータの MSE(ネットワーク 1).....	30
図 5.7 ch2 におけるテストデータの MSE(ネットワーク 1).....	30
図 5.8 ch3 におけるテストデータの MSE(ネットワーク 1).....	31
図 5.9 ch4 におけるテストデータの MSE(ネットワーク 1).....	31
図 5.10 ch1 における学習データの MAPE(ネットワーク 1).....	32
図 5.11 ch2 における学習データの MAPE(ネットワーク 1).....	32
図 5.12 ch3 における学習データの MAPE(ネットワーク 1).....	33
図 5.13 ch4 における学習データの MAPE(ネットワーク 1).....	33
図 5.14 ch1 におけるテストデータの MAPE(ネットワーク 1) .....	34
図 5.15 ch2 におけるテストデータの MAPE(ネットワーク 1) .....	34
図 5.16 ch3 におけるテストデータの MAPE(ネットワーク 1) .....	35
図 5.17 ch4 におけるテストデータの MAPE(ネットワーク 1) .....	35
図 5.18 ch1 における学習データの MSE(ネットワーク 2).....	37
図 5.19 ch2 における学習データの MSE(ネットワーク 2).....	37

図 5.20 ch3 における学習データの MSE(ネットワーク 2).....	38
図 5.21 ch4 における学習データの MSE(ネットワーク 2).....	38
図 5.22 ch1 におけるテストデータの MSE(ネットワーク 2).....	39
図 5.23 ch2 におけるテストデータの MSE(ネットワーク 2).....	39
図 5.24 ch3 におけるテストデータの MSE(ネットワーク 2).....	40
図 5.25 ch4 におけるテストデータの MSE(ネットワーク 2).....	40
図 5.26 ch1 における学習データの MAPE(ネットワーク 2).....	41
図 5.27 ch2 における学習データの MAPE(ネットワーク 2).....	41
図 5.28 ch3 における学習データの MAPE(ネットワーク 2).....	42
図 5.29 ch4 における学習データの MAPE(ネットワーク 2).....	42
図 5.30 ch1 におけるテストデータの MAPE(ネットワーク 2).....	43
図 5.31 ch2 におけるテストデータの MAPE(ネットワーク 2).....	43
図 5.32 ch3 におけるテストデータの MAPE(ネットワーク 2).....	44
図 5.33 ch4 におけるテストデータの MAPE(ネットワーク 2).....	44
図 5.34 ch1 における学習データの MSE(ネットワーク 3).....	46
図 5.35 ch2 における学習データの MSE(ネットワーク 3).....	46
図 5.36 ch3 における学習データの MSE(ネットワーク 3).....	47
図 5.37 ch4 における学習データの MSE(ネットワーク 3).....	47
図 5.38 ch1 におけるテストデータの MSE(ネットワーク 3).....	48
図 5.39 ch2 におけるテストデータの MSE(ネットワーク 3).....	48
図 5.40 ch3 におけるテストデータの MSE(ネットワーク 3).....	49
図 5.41 ch4 におけるテストデータの MSE(ネットワーク 3).....	49
図 5.42 ch1 における学習データの MAPE(ネットワーク 3).....	50
図 5.43 ch2 における学習データの MAPE(ネットワーク 3).....	50
図 5.44 ch3 における学習データの MAPE(ネットワーク 3).....	51
図 5.45 ch4 における学習データの MAPE(ネットワーク 3).....	51
図 5.46 ch1 におけるテストデータの MAPE(ネットワーク 3).....	52
図 5.47 ch2 におけるテストデータの MAPE(ネットワーク 3).....	52
図 5.48 ch3 におけるテストデータの MAPE(ネットワーク 3).....	53
図 5.49 ch4 におけるテストデータの MAPE(ネットワーク 3).....	53
図 5.50 km1-70 における楽曲ごとの比較 1.....	55
図 5.51 km1-70 における楽曲ごとの比較 2.....	55

図 5.52 km2-70 における楽曲ごとの比較 1 .....	56
図 5.53 km2-70 における楽曲ごとの比較 2 .....	56
図 5.54 km3-70 における楽曲ごとの比較 1 .....	57
図 5.55 km3-70 における楽曲ごとの比較 2 .....	57
図 5.56 km2-4 における楽曲ごとの比較 1 .....	58
図 5.57 km2-4 における楽曲ごとの比較 2 .....	58
図 5.58 km3-4 における楽曲ごとの比較 1 .....	59
図 5.59 km3-4 における楽曲ごとの比較 2 .....	59
図 5.60 km1-70 におけるジャンルごとの比較 1 .....	62
図 5.61 km1-70 におけるジャンルごとの比較 2 .....	62
図 5.62 km2-70 におけるジャンルごとの比較 1 .....	63
図 5.63 km2-70 におけるジャンルごとの比較 2 .....	63
図 5.64 km3-70 におけるジャンルごとの比較 1 .....	64
図 5.65 km3-70 におけるジャンルごとの比較 2 .....	64
図 5.66 km2-4 におけるジャンルごとの比較 1 .....	65
図 5.67 km2-4 におけるジャンルごとの比較 2 .....	65
図 5.68 km3-4 におけるジャンルごとの比較 1 .....	66
図 5.69 km3-4 におけるジャンルごとの比較 2 .....	66
図 5.70 km1-70 において T1 を複数回聴取した場合の比較.....	69
図 5.71 km1-70 において T7 を複数回聴取した場合の比較.....	69
図 5.72 km2-70 において T1 を複数回聴取した場合の比較.....	70
図 5.73 km2-70 において T7 を複数回聴取した場合の比較.....	70
図 5.74 km3-70 において T1 を複数回聴取した場合の比較.....	71
図 5.75 km3-70 において T7 を複数回聴取した場合の比較.....	71
図 5.76 km2-4 において T1 を複数回聴取した場合の比較.....	72
図 5.77 km2-4 において T7 を複数回聴取した場合の比較.....	72
図 5.78 km3-4 において T1 を複数回聴取した場合の比較.....	73
図 5.79 km3-4 において T7 を複数回聴取した場合の比較.....	73
図 5.80 km1-70 における 2 回目のクラスタリング結果 .....	75
図 5.81 km2-70 における 2 回目のクラスタリング結果 .....	75
図 5.82 km3-70 における 2 回目のクラスタリング結果 .....	75
図 5.83 km2-4 における 2 回目のクラスタリング結果 .....	76

図 5.84 km3-4 における 2 回目のクラスタリング結果 ..... 76

## 表一覧

表 3.1 実験に使用した楽曲 1.....	6
表 3.2 実験に使用した楽曲 2.....	7
表 3.3 実験に使用した楽曲 3.....	8
表 3.4 実験に使用した楽曲 4.....	9
表 3.5 実験で使用した評価指標.....	11
表 4.1 各印象評定語に対する SVM での平均分類精度.....	14
表 4.2 $w$ を変化させたときの SVM の平均分類精度.....	15
表 4.3 $w$ を変化させたときの SVM の平均分類精度( $n=5$ で固定).....	16
表 4.4 検証 5 におけるネットワーク構造.....	17
表 4.5 検証 6 におけるネットワーク構造.....	19
表 4.6 検証 7 におけるネットワーク構造.....	21
表 4.7 検証 8 におけるネットワーク構造.....	23
表 4.8 検証 8 における入力層、畳み込み層、プーリング層、LSTM 層の詳細.....	23
表 5.1 ネットワーク 1 の層構造.....	27
表 5.2 ネットワーク 2 の層構造.....	36
表 5.3 ネットワーク 3 の層構造.....	45
表 5.4 km1-70 で被験者に聴取させた音源.....	78
表 5.5 km2-70 で被験者に聴取させた音源.....	78
表 5.6 km3-70 で被験者に聴取させた音源.....	78
表 5.7 km2-4 で被験者に聴取させた音源.....	78
表 5.8 km3-4 で被験者に聴取させた音源.....	78

## 謝辞

3年間に渡ってご指導いただいた、亀山先生、菅沼先生に深くお礼申し上げます。また、オートエンコーダを利用する際に、多くのアドバイスをくださった Anh 先輩にもお礼申し上げます。最後になりますが、多くの助言をくださった亀山研究室のメンバーに感謝いたします。

## 参考文献

- [1] <http://www.apple.com/jp/music/> (2018/1/28 最終確認)
- [2] <https://music.line.me/> (2018/1/28 最終確認)
- [3] 毛利元昭, ”脳波計測実験のための簡易で安価な環境構築”, *Com=コム:愛知大学情報メディアセンター紀要* 26(1), 49-62, 2016-03-24
- [4] 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.
- [5] 小川宜洋, スティーブン・カルンガル, 満倉靖恵, 福見稔, 赤松則夫, ”ニューラルネットワークを用いた音楽聴取時の脳波解析”, *電子情報通信学会技術研究報告.NC, ニューロコンピューティング* 107(92), pp. 5-9, 2007年6月
- [6] 吉田真嵩, 菅沼睦, 亀山渉, ”基礎律動を用いた音楽聴取者の楽曲に対する印象推定に関する検討”, *電子情報通信学会 2016年全国大会*, 2016年3月15日
- [7] 澁田留奈, 菅沼睦, 亀山渉, ”基礎律動による楽曲聴取順序の識別に関する一検討”, *電子情報通信学会 2017年総合大会*, 2017年3月
- [8] 三木亮祐, ”基礎律動を用いた印象の類似性に基づく音楽分類に関する研究”, *早稲田大学基幹理工学研究科修士論文*, 2016年2月1日
- [9] <http://dev.choosemuse.com/hardware-firmware/headband-configuration-presets>  
(2018/1/28 最終確認)
- [10] <https://cran.r-project.org/web/packages/kernlab/kernlab.pdf>  
(2018/1/28 最終確認)
- [11] <https://www.csie.ntu.edu.tw/~cjlin/libsvm/> (2018/1/28 最終確認)
- [12] <https://keras.io/> (2018/1/28 最終確認)
- [13] <https://www.tensorflow.org/> (2018/1/28 最終確認)
- [14] 宇野毅明, 中原孝信, 前川浩基, 羽室行信, ”データ研磨によるクリーク列挙クラスタリング”, *情報処理学会研究報告アルゴリズム(AL)* 2014-AL-146(2), pp.1-8, 2014-01-23
- [15] kgocho, *DataPolishing*, 2016, Github repository ,  
<https://github.com/kgocho/DataPolishing> (2018/1/28 最終確認)
- [16] 吉田真嵩, 菅沼睦, 亀山渉, ”聴取楽曲と基礎律動の対応関係に関する一検討”, *HCGシンポジウム 2017*, 2017年12月



# 付録 学習を行った際に予測精度が十分でなかったオートエンコーダの層構造

表 A 学習を行った際の予測精度が低かったオートエンコーダ層の構造

チャンネル	周波数	層数(各層のノード数)	活性化関数	損失関数	エポック数	学習データ (最終エポック)	テストデータ (最終エポック)
AF7	1-45	14(45,100,400,500,500,500, 100,45,500,500,500,400,100,45)	relu	MAPE	1500	MAPE:64.79	MAPE:63.56
AF7	1-45	14(45,50,80,80,80,80,80,45, 80,80,80,80,80,45)	relu	MAPE	1500	MAPE:47.61	MAPE:47.85
AF7	1-45	15(45,50,80,80,80,80,80,80, 80,50,45,80,80,50,45)	relu	MAPE	1500	MAPE:48.55	MAPE:48.62
AF7	1-45	15(45,50,80,100,200,500,200, 100,80,50,45,100,80,50,45)	relu	MAPE	1500	MAPE:46.56	MAPE:47.33
AF7	1-45	15(45,50,80,100,200,500,200, 100,80,50,45,100,80,50,45)	relu	MSE	1500	MSE:4.13e-5 MAPE:33.74	MSE:6.52e-5 MAPE:33.25
AF7	1-45	21(45,50,80,100,200,500,800, 800,500,200,100,80,50,45,800, 800,500,100,80,50,45)	relu	MSE	1500	MSE:8.94e-5 MAPE:51.76	MSE:1.64e-4 MAPE:52.9
AF7	1-45	15(45,50,80,100,500,1000, 500,100,80,50,45,100,80,50,45)	relu	MSE	1500	MSE:5.13e-5 MAPE:37.63	MSE:8.27e-5 MAPE:41.22
AF7	1-45	21(45,50,80,100,500,1000,500, 100,80,50,45,50,80,100,500, 1000,500,100,80,50,45)	relu	MSE	1500	MSE:7.98e-5 MAPE:47.70	MSE:1.20e-4 MAPE:48.13
AF7	1-45	15(45,50,80,100,500,1000, 500,100,80,50,45,100,80,50,45)	relu	MSE	3000	MSE:5.25e-5 MAPE:35.74	MSE:7.56e-5 MAPE:36.26
AF7	1-45	17(45,50,80,100,500,100,80, 50,45,50,80,100,500,100, 80,50,45)	relu	MSE	1500	MSE:6.07e-5 MAPE:41.04	MSE:9.24e-5 MAPE:40.52
AF7	1-45	18(45,50,80,100,500,1000,500, 1000,500,100,80,50,45,500, 100,80,50,45)	relu	MSE	3000	MSE:6.66e-5 MAPE:40.86	MSE:9.92e-5 MAPE:41.23
AF7	1-45	22(45,50,80,100,500,1000,500, 1000,500,100,500,1000,500, 100,80,50,45,500,100,80,50,45)	relu	MSE	3000	MSE:1.20e-4 MAPE:59.11	MSE:1.74e-4 MAPE:58.36
AF7	1-45	20(45,50,80,100,500,1000,500, 1000,500,100,500,1000,500, 100,80,50,45,100,80,45)	relu	MSE	3000	MSE:1.08e-4 MAPE:56.57	MSE:1.59e-4 MAPE:56.40
AF7	1-45	14(45,50,80,100,500,1000,500, 100,80,50,45,100,80,50)	relu	MSE	3000	MSE:6.91e-4 MAPE:104.4	MSE:6.91e-4 MAPE:80.32

## 研究業績

題目	発表年月	発表	連名者
基礎律動を用いた音楽 聴取者の楽曲に対する 印象推定に関する検討	2016年3月	2016年 電子情報通信学会 全国大会	菅沼睦、亀山渉
聴取楽曲と基礎律動の 対応関係に関する一検討	2017年12月	HCG シンポジウム 2017	菅沼睦、亀山渉