

論文 / 著書情報  
Article / Book Information

題目(和文)	匂いセンシング空間の多次元データ解析による要素臭の探索法の研究
Title(English)	
著者(和文)	原田 祐希
Author(English)	Yuki Harada
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10146号, 授与年月日:2016年3月26日, 学位の種別:課程博士, 審査員:中本 高道,岩本 光正,中川 茂樹,佐藤 誠,間中 孝彰,竹井 義法
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10146号, Conferred date:2016/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

博士論文

「匂いセンシング空間の  
多次元データ解析による要素臭の探索法の研究」

原田 祐希

東京工業大学大学院 理工学研究科

電気電子工学専攻

2016年 3月

# 目次

第一章 序論	1
1.1. はじめに	1
1.2. 「匂い」の定義と全般的な嗅覚の仕組み	2
1.3. 嗅覚ディスプレイとバーチャルリアリティ	3
1.3.1. 嗅覚再現と「匂いの近似」について	3
1.3.2. 匂いの近似	3
1.4. 匂いの分類仮説	4
1.5. 官能ベースの匂いの分類	5
1.5.1. Hans Henning の匂いの分類	5
1.5.2. 岡安慎二らの日本の日常生活臭の分類	6
1.5.3. フレーバーホイール・フレグランスホイールについて	7
1.5.4. T&T オルファクトメーター	8
1.5.5. 官能ベースの匂いの分類まとめ	8
1.6. 嗅覚受容体発見以前の匂いの受容仮説	8
1.6.1. アムーアの立体受容説	9
1.6.2. 波動説	10
1.6.3. 分子振動説-ルカ・トゥリンの波動説	10
1.7. 生物の嗅覚受容体の仕組みと受容特性研究	11
1.7.1. バックとアクセル	11
1.7.2. 嗅覚受容体の働き	11
1.7.3. GPCR 受容体<Gタンパク質シグナル伝達系>	12
1.7.4. イオントロピック受容体	13
1.8. 匂いセンシング空間の定義	14
1.9. 匂いセンシングデータの取得方法	15
1.9.1. 感応膜や多孔質などによる表面修飾したガスセンサ(QCM、無機半導体、有機半導体センサ)	15
1.9.2. GC、LC(GC-MS 含む)	16
1.9.3.0. 検出器	18
1.9.3.1. FID (Flame Ionization Detector, 水素炎イオン化型検出器)	18
1.9.3.2. TCD (Thermal Conductivity Detector, 熱伝導度型検出器)	18
1.9.3.3. MSD (質量分析型検出器)	18
1.9.4. 官能評価を併用したガスクロマトグラフィー	18
1.9.5. MS(質量分析)	19
1.9.6. 官能、生理学的・分子生物学的手法	19
1.9.6.1. Dravnieks の多様な香気物質に対する官能評価データ	19
1.9.6.2. fMRI による脳の匂い応答	19

1.9.6.3. マウスの嗅球マップ .....	21
1.9.6.4. 昆虫の嗅球マップ .....	22
1.9.7. 嗅覚受容体を用いた匂いセンシングデータの取得方法.....	22
1.9.7.1. 嗅覚受容体の分子生物学と受容特性.....	22
1.9.7.2. 電気生理学的手法 .....	23
1.9.7.3. Hallem ら調査一香気物質への応答挙動.....	24
1.9.7.4. 指示薬利用法(染色法).....	24
1.9.7.5. 神崎研究室セルベースバイオセンサー .....	24
1.9.7.6. 蛍光プレートリーダー(FDSS)を用いた昆虫の嗅覚受容体応答測定について .....	25
1.9.7.7. ヒト嗅覚受容体による匂いセンシングデータ取得.....	26
1.10. 多次元データ解析を用いた解析 .....	27
1.10.1. Dravnieks の官能評価データに基づく検討 .....	27
1.10.1.1. Noam Sobel らの嗅覚におけるホワイトノイズの定義 .....	27
1.10.1.2. Jason B. Castro らの匂い認識空間のクラスター化解析 .....	28
1.10.1.3. 匂いの認識空間の主成分分析と匂いへの嗜好性の関係.....	29
1.10.1.4. RD Newcomb によるスーパーEnose の提案.....	31
1.10.1.5. ドイツのリューベック大学ママルーク .....	33
1.10.2. マウスの嗅球マップデータに基づく検討.....	34
1.10.2.1. 嗅球活動の画像に基づく初段の嗅覚系の能力の分析 .....	34
1.10.3. 既存研究 ; 各種センサ(MS, QCM etc) を用いた近似臭作成方法の検討.....	36
1.10.3.1. これまでの近似臭作成方法 .....	36
1.10.3.2. 近似臭と対象臭の高 m/z 領域におけるマススペクトルの類似性 .....	36
1.11. 本研究の目的と論文構成 .....	38
第二章 非負値行列因子分解法を用いた近似臭作成方法における距離尺度の検討.....	43
2.1 データベースとその基底について.....	44
2.1.1 主成分分析 (PCA) .....	44
2.1.2 非負値行列因子分解法 (NMF) .....	45
2.2 マススペクトルのデータにもとづく NMF 法を用いた要素臭の探索.....	45
2.2.1 マススペクトルのデータの特徴と本研究で使用したデータ .....	45
2.2.2 マススペクトルのデータ取得方法 .....	46
2.2.3 中本研究室での質量分析(MS)による要素臭の作成と近似臭の作成の検討 .....	47
2.3 NMF 法における距離尺度 .....	48
2.4 近似レンピ計算における IS 距離の導入と MATLAB 上実装.....	50
2.5 異なる距離尺度に基づく NMF 法を適用して得た基底ベクトルを用い、ターゲット香気を近似した結果 .....	52
2.6 まとめ .....	56

第三章 嗅覚受容体応答情報に基づいた非負値行列分解法による要素臭の探索.....	57
3.1 NMF 法適用のための応答データの非負化.....	59
3.2 NMF 法における基底数.....	60
3.3 基底行列 W の評価方法と行列因子分解の条件.....	61
3.4 行列因子分解で求められた結果の評価.....	63
3.5 10 種類の要素臭を使用した匂い応答パターンの再現.....	66
3.6 NMF 法を用いた要素集作成方法 ; 匂い近似における要素臭の数と近似臭の精度の相関性の検討 .....	67
3.7 まとめ.....	68
第四章 香気分子構造の数値表現と様々な香気分子によるパラメータ空間.....	69
4.1 嗅覚受容体の応答強度予測の意義.....	69
4.2 パラメータ空間と SOM について.....	70
4.3 香気物質の分子構造空間の解析.....	72
4.3.1 MATLAB における基本的な SOM の実装.....	72
4.4 化合物の構造表記方法と分子構造パラメータ.....	73
4.4.1 コンピューター上で取り扱いやすい化合物の構造表記方法.....	73
4.4.2 嗅覚受容体の応答強度予測に用いる分子構造情報.....	75
4.5 パラメータ空間と SOM.....	77
4.5.1 香気物質の空間に基づく、パラメータ空間の非線形マッピング.....	77
4.5.2 化合物の構造パラメータ空間の非線形マッピング.....	79
4.6 まとめ.....	81
第五章 香気分子構造をもとにした自己組織化写像による嗅覚受容体の応答強度予測手法.....	82
5.1 SOM による非線形の回帰予測.....	82
5.1.1 パラメータ空間の非線形マッピングを利用した予測手法.....	82
5.2 6 種類の ORN (Or22a, Or35a, Or67a, Or47b, Or65a, Or88a) の 108 種類の香気物質に対する応答 強度.....	83
5.2.1 自己組織化写像(SOM)による嗅覚受容体の応答強度予測.....	83
5.3 SOM サイズ.....	84
5.4 変数(分子情報)選択法.....	85
5.4.1 条件数に基づく選定.....	86
5.4.2 使用するパラメータ個数と条件数の関係.....	88
5.4.3 ‘AIC’ を用いたパラメータ選択(使用するパラメータ数)の検討.....	88
5.4.4 交差検定の AIC による評価に基づく最適なパラメータの選定.....	90
5.4.5 提案する最適各パラメータ組合せ探索法.....	92
5.5 最適各パラメータ組合せ探索法を用いた自己組織化写像(SOM)による嗅覚受容体の応答強度予測 .....	93

5.5.1 電気生理学的測定の結果に対する自己組織化写像(SOM)による嗅覚受容体の応答強度予測	93
5.5.2 香気物質の空間に基づく、能動学習のための検討	95
5.6 主成分分析 (PCA) による香気分子の構造パラメータ圧縮とそれを用いた予測	99
5.6.1 主成分分析 (PCA) による香気分子の構造パラメータ圧縮	99
5.6.2 主成分分析 (PCA) と変数選択を使用した SOM による ORN 応答予測	100
5.7 蛍光プレートリーダーを用いた昆虫の嗅覚受容体応答測定の影響強度予測	103
5.8 SOM による応答予測における 3 値化検討	108
5.9 まとめ	109
第六章 香気分子構造情報のオートエンコーダーによる次元圧縮と多層パーセプトロンによる嗅覚受容体の応答強度予測手法	111
6.1 オートエンコーダー(砂時計型ニューラルネットワーク)	111
6.2 分子構造パラメータ空間の次元圧縮	112
6.3 香気分子の分子構造パラメータへのオートエンコーダー(砂時計型ニューラルネットワークの適用)	113
6.3.1 データについて	113
6.3.2 オートエンコーダーのバックプロパゲーション学習手順	114
6.3.3 中間層のユニット数	114
6.3.4 バックプロパゲーション学習の方法	115
6.4 SOM を用いた応答予測	119
6.4.1 オートエンコーダーによる香気化合物の分子パラメータを圧縮と SOM を用いた応答予測(93 化合物～OR13a)	119
6.4.2 SOM を用いた応答予測における 3 値化の効果	120
6.5 オートエンコーダーによるパラメータ空間の圧縮をもとにしたニューラルネットワークへのバックプロパゲーション学習による予測(OR13a)	122
6.5.1 オートエンコーダーによる香気化合物の分子パラメータを圧縮とニューラルネットワークを用いた応答予測(93 化合物～OR13a)	122
6.5.2 オートエンコーダーによる全般的な香気化合物の分子パラメータを圧縮	124
6.5.3 香気化合物の分子パラメータの全般的なオートエンコーダー圧縮に基づくニューラルネットワークを用いた応答予測(93 化合物～OR13a)	125
6.6 まとめ	125
第七章 結論と今後の課題	127
謝辞	129
参考文献	130
学会発表など	135



# 第一章 序論

## 1.1. はじめに

特に近年、匂いを提示する嗅覚ディスプレイの開発は関心を集めており、映像や音と共にバーチャルリアリティー再現の一環として香りをユーザーに提示するなど、様々な分野で応用が期待されている。嗅覚ディスプレイなどを用いることで、複数の香りを任意の比率で混ぜ合わせて調合し、すぐに香りを発生させることができる。しかし、香りの汎用的な要素臭はまだ明らかになっておらず、開発が期待されてきた。

匂いは香気物質の混合物であり、生体内で流通する匂い情報は複合的な嗅覚神経細胞(ORN)応答情報に象徴されるような"複合的な信号"である。香気物質を含む空気は鼻腔から入り、嗅絨毛上に存在する ORN において香気物質～受容体相互作用が起こり、神経信号が発される。一般に生物は受容挙動特性の異なる複数の ORN を持っており、その性質の差異は神経細胞上に存在する嗅覚受容タンパク質と香気分子の相互作用の差異に依存する。複数種類のおい受容体によって発生した神経信号は脳へと伝わり、匂いの知覚・認知が起こる。

筆者は香気物質への生体応答を多次元データ解析して、要素臭の計算手法を確立する必要があると考えた。加えて、受容体の機能解明が実験で用いられた香気物質に限定されており、機械学習的なアプローチを用いて受容の予測を行うことに興味を持った。

東京工業大学・中本研究室では、自由な遠隔地での匂い再現を目指して、匂いの記録再生システム、嗅覚ディスプレイについて研究を行ってきた。多次元データ解析的な手法に基づき、香りの科学の成果を当研究室から香りの技術として還元したいと考えている。

## 1.2. 「匂い」の定義と全般的な嗅覚の仕組み

匂いとは、人体において揮発性化合物の混合物が鼻の粘膜上の嗅覚神経細胞にて受容され、それによって生じる感覚刺激の対象である。そのような感覚刺激をもたらす化学物質を匂い物質、あるいは香気分子と呼ぶことにする。匂い物質はフレーバー、フレグランス、アロマテラピー・芳香療法などにおいて、何世紀にもわたって治療目的や食品添加物のために利用されており、人類と密接にかかわっている。[001]

香気物質として 40 万種類もの物質も存在することが知られている。全般的な嗅覚の仕組みは図 1-01 に示すように、科学的に理解されている。鼻腔から取り込まれた匂い物質は、嗅繊毛上に存在する ORN (olfactory receptor neuron; 嗅覚神経細胞)において、香気物質～受容体相互作用が起こる。このとき、香気物質～受容体相互作用を起こっており、生物の持つ嗅覚受容神経細胞の応答特性の差異は、細胞膜上に存在する嗅覚受容タンパク質と香気分子の相互作用の差異に依存すると考えられている。この結果 ORN から神経信号が発される(嗅覚神経細胞(ORNs)信号)。ほ乳類および昆虫では ORN 一個当たり単一の受容体を膜上に発現する。生物は数十～数百種類の ORN を持つ。複数種類のおい受容体によって発生した複合的なセンシング情報は脳に送られる。神経信号は脳へと伝わり、刺激情報は統合され、脳内嗅覚地図との照合される結果、匂いの知覚・認知が起こる。五感のうち、最も早く様々な動物の祖先に出現したものは化学感覚であるとされており、化学感覚は外界に無数に存在する化学物質の混合物を、その ORNs 信号パターンとして、離散的情報として処理する。[002]

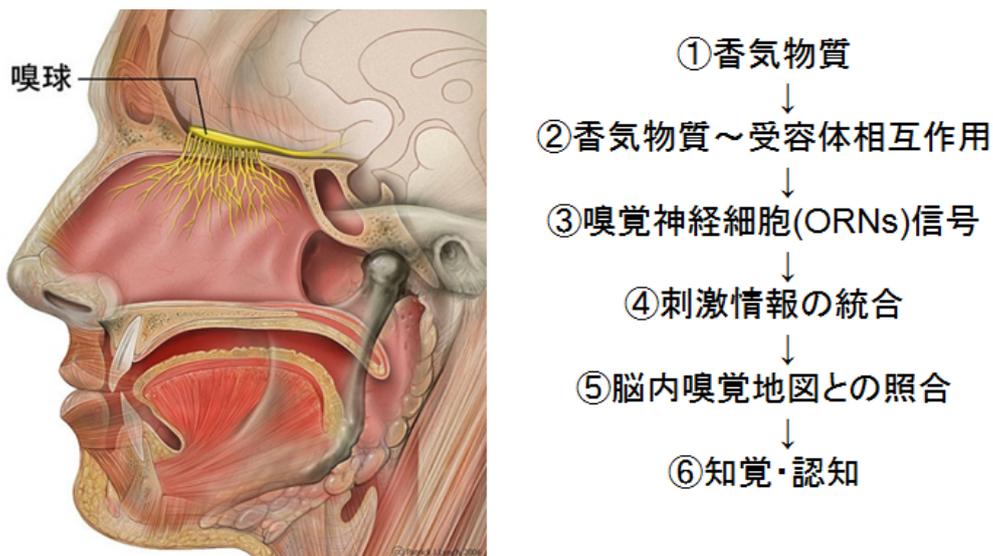


図 1-01. 全般的な嗅覚の仕組み [002]

### 1.3. 嗅覚ディスプレイとバーチャルリアリティ

#### 1.3.1. 嗅覚再現と「匂いの近似」について

図 1-02 は電磁弁型嗅覚ディスプレイである。この装置は高速電磁弁開閉方式を用いたもので、装置の内部にセットした各匂いサンプルの匂い濃度を電磁弁の開閉頻度で制御することが可能である。異なる香りを混ぜて嗅ぐと異なるイメージの香りが作り出せる。この装置を使えば、リアルタイムに気相状態で匂い物質を調合し提示することが可能であるため、対象の匂いの情報をもとに近似した匂いを作る技術が求められる。



**図 1-02. 嗅覚ディスプレイとバーチャルリアリティ [003]**

#### 1.3.2. 匂いの近似

数種類の香りを混ぜて嗅ぐと、異なるイメージの香りが作り出せる。匂いの情報を元に匂いを自由に遠隔地で再現する装置を実現するためには、図 1-03 に示すように複数の匂いを用意し、調合して目的の匂いを近似して再現できると考えられる。ただし、どんな香りでも再現ができるような汎用的な原臭については、諸説(後述)が提案されてきたが明らかになっていない。

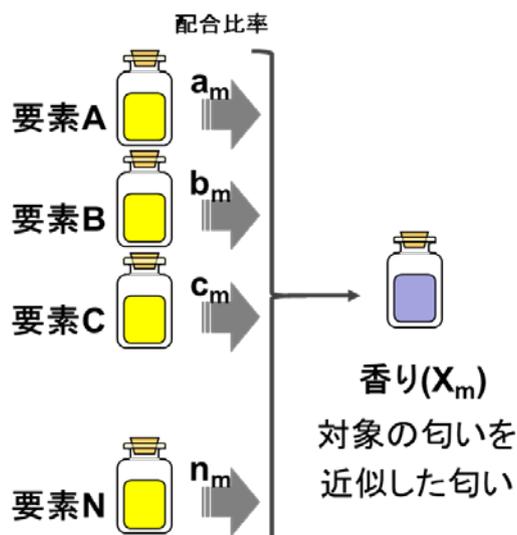


図 1-03. 嗅覚匂いの近似

#### 1.4. 匂いの分類仮説

現在の嗅覚における匂い受容機構に関する生物学的な理解は、1991 年の バックとアクセル: 匂い受容体発見の報告[004]以降の目覚ましい進歩によるものである。その科学的理解が受け入れられたことを象徴するのは 2004 年のノーベル賞受賞である。裏を返せば 21 世紀にはいる前後くらいまで、匂いの正体、匂い受容機構への理解には、いくつもの仮説があり、そのどれも決定力に欠けるものであった。まず古典的な匂いの受容仮説と匂いの系統分類案・要素臭案についてまとめ、バックとアクセル: 匂い受容体発見の報告以降の生物学的な理解は、1.6 節で言及する。匂いの正体、匂い受容機構へのバックとアクセル以前までの仮説は、1.7 節で言及する。

#### 年表

- 1756 リンネ: 匂いの分類
- 1860 フェヒナー: 精神物理学を創案
- 1895 ツワーデマーカー: 「匂いの生理学」を出
- 1916 ヘニング: 「嗅覚」を出版
- 1924 ヘニング: 匂いのプリズム
- 1938 ダイسن: "匂いの科学的基礎"を発表
- 1954 デイビーズ&テイラー: 孔あけ説
- 1956 オットソン: 嗅電図の応答記録
- 1963 ゲストランドら: 嗅細胞単独ユニットの応答記録
- 1964 ライト: 「匂いの科学」を出版; 振動説
- 1967 モンクリフ: 「化学感覚概説」を出版
- 1970 アムーア: 立体化学説
- 1975 スティーブンス: 知覚強度のベキ法則を提案
- 1977 ライト: 「匂いと分子振動」を発表
- 1991 バックとアクセル: 匂い受容体発見 (cell)

1995 バックとアクセル:”匂い分子の理論”を公表

1996 トウリン:”一次嗅覚受容の分光的機構”を公表 (chemical sense)

2004 バックとアクセル:ノーベル賞受賞

### 1.5. 官能ベースの匂いの分類

香りの分類は昔から多くの研究者によって試みられてきた。表 1-04 にこれまで提案されてきた分類を転記する。表中、横に官能的に同系統と思われる研究者らによって提案された原臭名を並べた。-は研究者による言及がないことを表す。表内の左端の「匂いのタイプ」には、調香トレーニングや香料の評価に使用される官能評価用語で最も近い名称を示した。以下の節で代表的な分類に関して説明をする。

表 1-04. これまでの研究者による原臭の分類 [006]

(匂いのタイプ)	linnus 1752	zwaardemaker 1895	henning 1916	crocker 1928	amoore 1962	schutz 1964	wright&michels 1964
Aroma and/or green	アロマ	アロマ	-	-	-	-	-
Ether. fruits	-	エーテル	果実	-	エーテル	エーテル	-
Heavy floral	芳香	芳香	-	-	-	芳香	-
Light floral	-	花	花	芳香	花	-	シト랄
Sweet	-	-	-	-	-	甘	-
Mint	-	-	-	-	ミント	-	-
Resinoic	-	-	樹脂	-	-	-	樹脂
Spice or pepper	-	-	薬味	-	樟腦	薬味	薬味
Amber/musk	アンバー	アンバー	-	-	ムスク	-	-
Burnt	-	焦	焦燃	焦燃	-	焦燃	-
Garlic or onion	ニンニク	ニンニク	-	-	-	-	-
Caproic	山羊	山羊	-	動物	-	-	-
Light acidic	-	-	-	酸	刺激臭	-	(三叉神経刺激)
Sulphur (putrid)	腐敗	嫌悪	腐敗	-	腐敗	硫黄	不快
Foul	催嘔	糞臭	-	-	-	-	ベンゾチアゾール
Rancid	-	-	-	-	-	rancid	-
Oil	-	-	-	-	-	油膜	-
Metallic	-	-	-	-	-	金属性	-

#### 1.5.1. Hans Henning の匂いの分類

Hans Henning は“Odor Prism”として、ドイツの Henning が匂いの体系化を試みた際に薬味、花、果実、樹脂、焦げ臭、腐敗の6臭を代表的な匂いとして挙げた。各香気の類似度や近さを考慮して匂いの各要素のイメージを3次元内に配置すると、Henning は図 1-05 に示すようなプリズム(三角柱)の各頂点に各要素が位置し、現実の香りもプリズム内にマッピングできると考えた。

1. Fragrant 花(ラベンダーやバラの花のにおい)
2. Etheral エーテル臭(エーテルや洗剤のようなおい)
3. Resinous 樹脂臭(樹脂やテレペンチンのようなおい)
4. Spicy 薬味臭(シナモンやナツメグのようなおい)
5. Putid 腐敗臭(糞や腐った卵のようなおい)
6. Burned 焦臭(タールのようなおい)



合が小さく、食品や生活用品としての化学物質に関する匂いが大きい位置を占めたという。[008]

### 1.5.3. フレーバーホイール・フレグランスホイールについて

フレーバーホイール・フレグランスホイールは官能評価において頻繁に目にし、実際に用いられている。ビール、ワイン、ウイスキー、日本酒、もしくはチョコレート、和風だし、コーヒーなど、幾つものフレーバーホイールの資料がネット上で見られる。ホイールを使う方法は官能評価と商品開発においてはとても有用であるが、それは要素臭を考えたり香気物質の香調に与える影響を見積もったりするのは異なる。これらは食品・農産品・工業製品の匂いの分類や、フレーバー、フレグランスの分類に用いられる。例として図 1-07 にワインのアロマホイールとフレグランスホイールを示す。

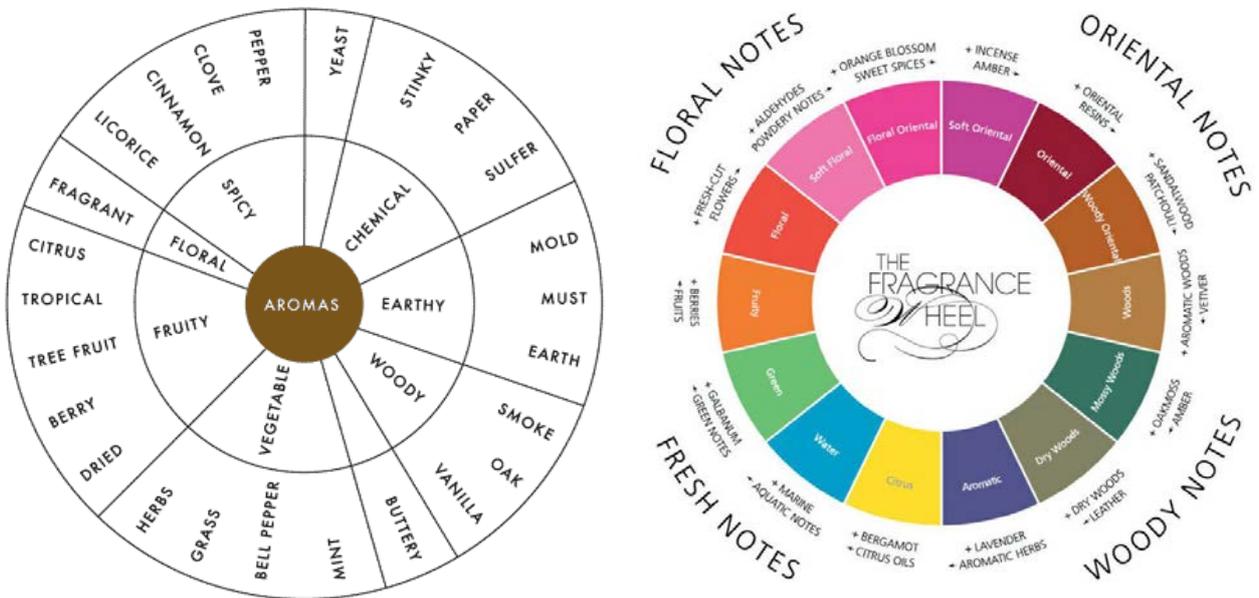


図 1-07. (A)ワインのアロマホイールと(B)フレグランスホイール [009]

すでにその分野に関して評価経験を積んでいる商品開発などに携わるエバリュエーターが、その方向性について打ち出したもの、これがフレーバーホイール・フレグランスホイールである。フレーバーホイールはベテランエバリュエーターが都度作成するものであり、一種の言語による文化伝承的な側面がある。

盛り込む要素や大分類～小項目は、長年の経験を持つエバリュエーターや分析の技術者のディスカッションによって決定される。十分に経験を積んだ人間による、「評価項目に何を置いたら良いのか」がまとめられたフレーバーホイール・フレグランスホイールを共有知識として持つことで、食品・嗜好品開発における官能評価と試作のコミュニケーションが円滑になり正確になる。商品開発において官能評価と試作の繰返しを効率的にすることができる。

### 1.5.4. T&T オルファクトメーター

T&T オルファクトメーターは、わが国で唯一基準嗅覚検査として適用された、嗅覚検査キットである。図 1-08 に示す写真はキットの外観とにおいの成分と使用される希釈倍率の表である。体内診断薬として保険点数が定められ、嗅覚障害の程度や治療効果の判定、労災の補償判定のため、耳鼻咽喉科領域で広く使用されている。この 5 つの基準臭と濃度は医薬診断用の基準として策定しされている。

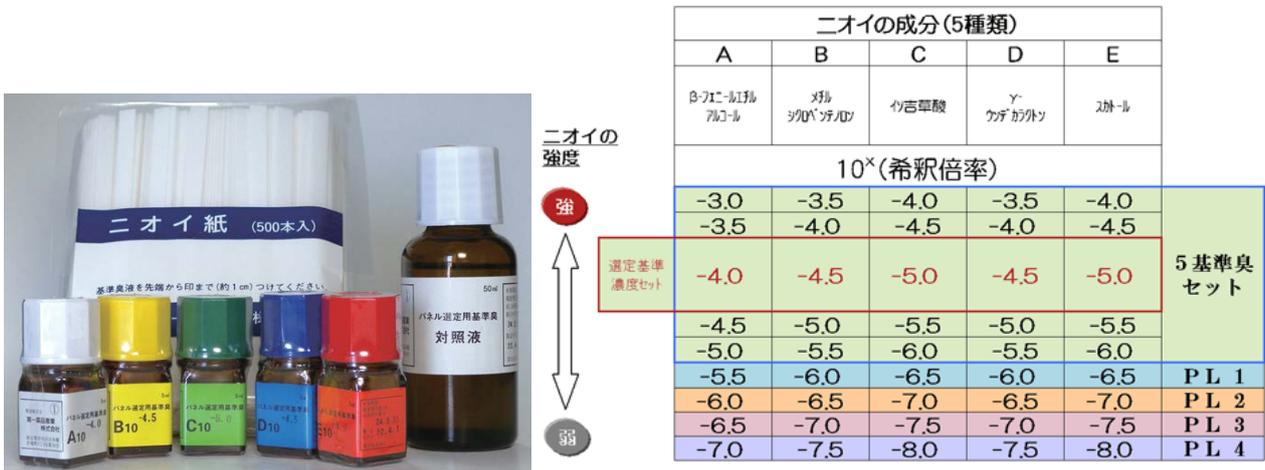


図 1-08. T&T オルファクトメーター [010]

香りの分類と表現方法、嗅覚機構に関する仮説と要素臭・臭素、匂いの官能面からの要素化、香りの受容メカニズムの解明について検討された上で、基準臭とその濃度について決定された。[010]

### 1.5.5. 官能ベースの匂いの分類まとめ

特にフレーバーホイール・フレグランスホイールについて、香気の官能評価に用いる天然の香気や製品の匂いを分類したものであり、要素臭に直接利用可能なものではない。また、例えば同じアニマルノートであっても、ジャスミンのアニマルノートはインドール、ローズのアニマルノートはフェニルアセテートであるように、要素名が同じであっても、フレーバーホイール・フレグランスホイールが異なると同じ要素名でも異なる香調である場合がある。要素臭に関する考察は、官能評価ベースの匂い分類やフレーバーホイールのような匂い要素の解釈とは異なる視点に基づく必要がある。

### 1.6. 嗅覚受容体発見以前の匂いの受容仮説

嗅覚に関するブレークスルーは、1990 年代のバックとアクセルによる嗅覚受容体に関する遺伝子が見出されたことである[004]。これによって彼らはノーベル賞を受賞し、嗅覚の“生化学的な検出機構”に一定の結論が与えられ、研究手法や考え方が一新された。ここでは吉儀の著書[006]を参考にバックとアクセル以前までに提唱された、近代科学の発展の中において匂い論について主要なものを説明する。

#### 1) 振動説(放射説)

…匂いを出す物質からそれを感じずる人まで、その間、光や音のような振動として伝わり、嗅上皮を刺激するという

説である。

## 2) 化学説

…匂いは、物質から四散した匂い分子が人の鼻腔内に飛び込み、嗅細胞の受容膜との間で化学反応を起こして刺激するという説、また化学反応ではなく吸着、脱着という物理化学的反應を起こして刺激とする相界説もこれに含まれる。立体受容説もこれに含まれる。

## 3) 酵素説

…刺激受容器の種類は単一～少数であり、匂い物質による嗅神経の応答の多様性が酵素による説であった。刺激受容器の表面にある酵素系に匂い物質が与える影響によって、匂いの相違が生じるとする説である。

## 4) 立体受容説

…別名「鍵と鍵穴」説で、匂いの違いは匂い分子の立体構造、つまり外形と大きさによって決まるとする説である。

### 1.6.1. アムーアの立体受容説

振動説を別にすれば、匂いは有機物質が揮発した匂い分子が人の鼻腔内に飛び込み、嗅覚神経を刺激起こる感覚だと考えられてきた。歴史をさかのぼって匂いに関する記述を探してみると、古代ローマの哲学者であるルクレチウスが、「匂いというものは、原子のつながった分子でできていて、分子の形や大きさの違いが、匂いの質の違いを生む」と記述している。近代科学の発展の中においては、Moncrieff(1951) [011]の考えを Amoore(1962) [012]が発展させた。アムーアの受容体説では、匂いの分子は外形と電荷によって 7 種類の基本臭に分けられる。それらは、樟脳臭、エーテル臭、ハッカ臭、麝香臭、花香、刺激臭、腐敗臭で最後の 2 つを除いて他の匂いは図 1-09 に示すような嗅細胞の受容器にある特殊なくぼみにそれぞれ嵌まり込むような外形をしているとした。

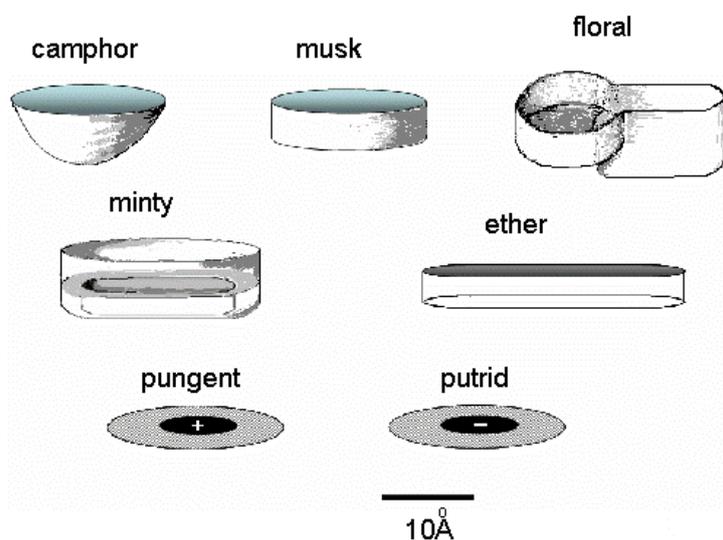


図 1-09. Amoore's went on to propose 7 primary odours [012]

アムーアの受容説は、バックとアクセルによって報告された細胞膜上の受容体タンパク質による嗅覚機構に最も近いものである。アムーアが受容説を提案した当時、当時受容体が発見されなかったため、立体受容説は決定的な効力を発揮しなかった。

### 1.6.2. 波動説

アムーアの立体受容説の同時期、多くの支持者を持っていたのが波動説である。匂いとはなんらかの波動であって、それが人間によって知覚されているのだというものである。

匂いの波動の本質についても様々な説が存在し、時代の変遷、科学の発展に従い様々に推測された。分子振動説とは、匂いの分子を構成する原子と原子の結合から生まれる特有の振動スペクトルを鼻のなかの組織、のちには嗅神経細胞が読み取っているという説であり、古くはマルコム・ダイソンやロバート・ライトなどが主張していた。分析化学の発展にしたがってある程度香気物質の存在が明らかになってきても、「形」だけでは説明がつかない匂いの不思議を説明するためにルカ・トゥリンのように波動説を発展させた分子振動説を嗅覚知覚の原理として主張・イメージする人も多かった。

### 1.6.3. 分子振動説-ルカ・トゥリンの波動説

ルカ・トゥリンは現在入手可能な香水を網羅した香水評価をまとめた著書[013]も出版すると同時に、振動論を支持していた研究者[014] [015]である。ルカ・トゥリンの「振動説」は分子振動説であり、嗅覚細胞上にはMS(質量分析器)やIR(赤外分光器)のような“振動数”を分析するような感覚があるのではないかとし、匂いと振動数の関連性検討、IR と香調の関連性を調査した。この理論を支持する一つの証拠として、これまでの(他の研究者による)研究で、ムスク化合物シクロペンタノンのヒト被験者が水素/重水素アイソトポマー(同位体原子で置換した異性体)の識別が起こると、報告されている。

図 1-10 にトゥリンの匂い受容機構の説の概要を示す。嗅覚受容体のポリペプチド鎖から形成されたバレル構造が存在している。トゥリンは匂いの受容が振動励起と連動していると考えし、匂いの受容体が非弾性トンネル分光の機構を持つとする仮説を発表した。非弾性トンネル分光法とは、試料分子(吸着分子など)を含む絶縁層をトンネルする電子と分子の相互作用を利用して、試料の振動励起スペクトルを得る方法である[014], [016]。受容体には二つのサイト、非弾性トンネル現象を発生させられる電子が存在するサイト RD とトンネリング後のサイト RA、が存在している。香り分子は、受容体に取り込まれて受容体を変形、香り分子の振動励起によって非弾性電子トンネルが図中 D と A との間に発生するとされている。その後香り分子が排出され、システムを再初期化されるとした。分子内水素原子を重水素化すると、化合物の軌道エネルギーなどは変わらないのだが、重水素原子の関わる IR 共振数がシフトし、分子振動数が変化する。この報告は分子振動と匂いの質との間に関係があることを示す証拠であるとされた。嗅覚の振動論は、嗅覚受容体(OR)の活性部位で、電子移動が匂い物質を横切って発生することを前提としており、匂い物質の振動数の敏感な指標となる上に、最終的に嗅覚知覚へとつながるとしている。

だが波動の受容器官は解明されず、本論は仮説に過ぎなかった。その後、バックとアクセルが嗅覚受容体の存在を証明し、それによりノーベル賞を受賞するなど立体受容説が科学的に確立した。

近年ヒト嗅覚 GPCR 研究者から、異種または発現系を用いた時に、シクロペンタムスコン明確な応答性を持つことが確認されているヒトムスク認識受容体、OR5AN1 を用いて、in vitro でのこれらの化合物のアイソトポマーの区別

に失敗することを見いだした。これは“分子内水素原子を重水素化すると、化合物の電気的物性は変わらないが、分子振動数が変化するため匂いの質が変化する”とした分子振動説にとっては不自然な実験結果である。さらに、ムスク受容体 OR5AN1 の結果と同様に、マウスの(メチルチオ)メタンチオール認識受容体、MOR244-3、ならびに他のヒトおよびマウスの OR、を用いて実験をしたところ、未重水素化体、重水素化体、およびそれらの 13C アイソトポマー、それぞれのリガンド対し反応には有意差がみられなかった。ここからも同様にこの分子振動説にとっては不自然な実験結果であるとされる。[017]

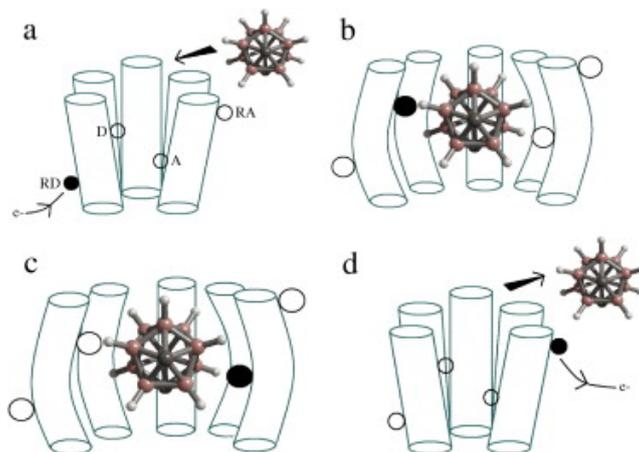


図 1-10. トウリンの匂い受容機構の説の概要 [014]

以上のように科学的反証が示されており、振動論を積極的に肯定できる科学根拠は今となっては無い。ヒト被験者によるシクロペンタノンの水素/重水素アイソトポマーの識別に関しても、ごく微量の不純物により匂いの印象が変化してしまうことが起きやすい再現性を確保することが難しい実験である。筆者としても、ヒト嗅覚がバックとアクセルにより提案された GPCR 型受容体(後述)による化学刺激受容であると理解している。

## 1.7. 生物の嗅覚受容体の仕組みと受容特性研究

### 1.7.1. バックとアクセル

1991 年の バックとアクセル: 匂い受容体発見の報告[004]以降の嗅覚における匂い受容機構に関する生物学的な理解の目覚ましい進歩について、これ以降の章で紹介する。

### 1.7.2. 嗅覚受容体の働き

脊椎動物や昆虫は、匂いを検出し、区別するために化学感覚受容体の複雑なレパートリーを進化させてきた。

哺乳動物において、嗅覚系は、数百～数千の嗅細胞に発現する嗅覚受容体(OR)が、多種多様な匂い分子の検出および識別における重要な役割を果たしている。匂い物質は、組み合わせ識別(combinatorial fashion)、単一の匂い物質は、受容体の特定の組み合わせを活性化し、それが特定の香りとしての認識につながるという方法で OR によって認識される。いくつかの例外を除いて、脊椎動物の化学感覚受容体は、G タンパク質(ヘテロ三量

体グアニンヌクレオチド〜結合タンパク質) 共役受容体(G-protein-coupled receptor、GPCR)のファミリーに属する。GPCR は細胞のシグナル伝達事象のカスケードを開始させ、最終的には電氣的にニューロンを励起する機能を持つ。GPCR は細胞膜中に埋め込まれた状態で機能し、7回タンパク質鎖が細胞膜を貫通する構造(7回膜貫通型構造)をもち、すべての GPCR に共通する構造となっている。

昆虫の受容体は構造的および遺傳的に脊椎動物の受容体とは無関係である。昆虫の受容体たんぱく質は二つのたんぱく質の複合体という構造を持ち、匂い物質の結合に応じてゲート(が開閉)するイオンチャンネル型受容体として動作する。またアミノ酸配列は脊椎動物の GPCR 型嗅覚受容体の配列とは異なる[018], [019]。神経細胞が受容体への匂い物質の結合により神経信号を生成する際、脊椎動物では信号の増幅や減衰に相当する機能が存在すると考えられる一方で、昆虫では神経信号生成までの処理速度を向上させているといわれている。以下にイオトロピック受容体と GPCR 受容体について説明する。[020]

### 1.7.3 GPCR 受容体<Gタンパク質シグナル伝達系>

嗅覚受容体(OR)のゲノムファミリーは、脊椎動物のゲノムにおける最大であり、脊椎動物の嗅覚をになっており、伝統的に 17 種類の OR サブファミリーに細分される。

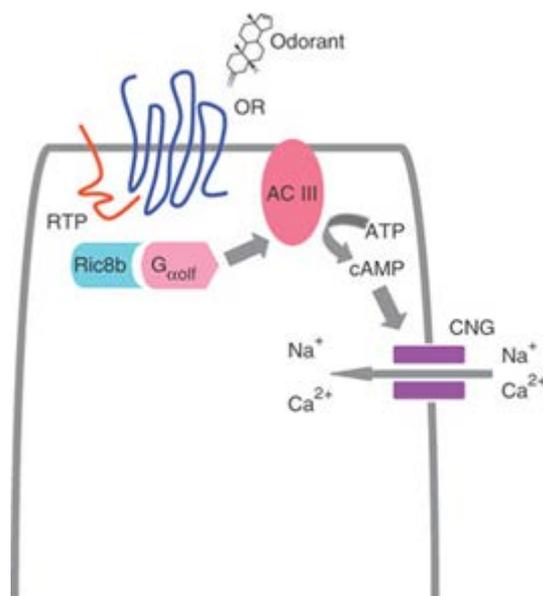


図 1-11. 化学物質が受容体と結合することで引き起こされるGタンパク質シグナル伝達 [021]

受容体と結合した化学物質の情報はGタンパク質シグナル伝達系を介して、高次中枢へ情報伝達される。化学物質が受容体と結合することで引き起こされるGタンパク質シグナル伝達の機構を図 1-11 に示す。匂い物質(odorant)と結合した嗅覚受容体(OR)の構造変化は、下流のGタンパク質(G<sub>olf</sub>)を活性化し、細胞内セカンドメッセンジャー濃度(cAMP)を上昇させる。セカンドメッセンジャーは環状ヌクレオチド作動性イオンチャンネル(CNG)を開口させ、細胞内に Ca<sup>2+</sup>, Na<sup>+</sup>カチオンが流入し、その結果、嗅神経細胞が脱分極する。

この OR は立体化学構造に対して優れた選択性を有し、分子構造の差異を識別できる。鏡像異性体は、同一の化学特性を示すにもかかわらず、異なる香りを発することがよくあることとして知られている。この差異を生む分子挙動を解明するために、鏡像異性体の関係である *l*-および *d*-メントールへのマウスの嗅球における応答を記録したところ、類似しているが重複した別個の受容体の活性化パターンを誘発したことを東原らは報告している。彼らは *l*-メントール/*d*-メントールへの選択性を示し応答する受容体を同定し詳細なその構造活性相関研究をすることで、鏡像異性体に対して特異的なメントール受容体の高い立体選択性を明らかにした。各 OR の結合部位は、非常によく似た匂い物質の構造の微妙な違いを区別するために進化したと考えられる。[022]-[033]

線虫の嗅覚システムはほ乳類とは異なるが、様々な匂い物質を感知し、応答する。嗅覚受容体は、GPCR であり、直接揮発性または水溶性の匂い物質と結合する。線虫は、嗅覚受容体や感覚神経レベルで感知でき、濃度を感知できる機構を持っていることが分かった。[022]-[033]

ヒトの嗅覚受容体群に関しても、約 400 種類の欠損しておらず発現する嗅覚受容体遺伝子が確認されている。松波らはヘテロ発現系アッセイを使用して、嗅覚受容体の機能を決定し、27 種類の嗅覚受容体について香気物質への応答挙動を報告した(後述) [034]

#### 1.7.4. イオントロピック受容体

ゲノムから得られた情報からは、昆虫の嗅覚シグナル伝達におけるGタンパク質の関与は示唆されなかった。嗅覚受容体の下流にGタンパク質が存在しないことが報告されている[005], [035]。昆虫の嗅覚トランスダクションの分子基盤は、図 1-12 に示すような嗅覚受容体複合体が活性化するGタンパク質非依存的なシグナル伝達経路が、その本質であると考えられた[036]。

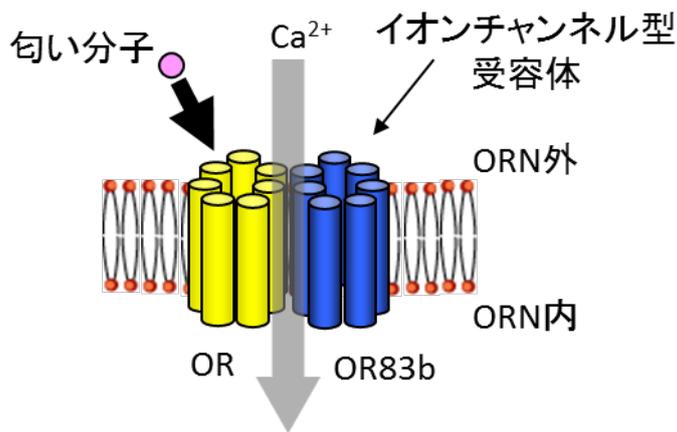


図 1-12. 昆虫の嗅覚[005]

OR83b ファミリー受容体は異なる昆虫種間で 70%以上の高い遺伝子配列の相同性を持ち、全ての昆虫で高度に保存され、別の受容体たんぱく質と複合体を構成することがわかった。[037], [024] OR83b は嗅覚受容体と複合体を構成することで、受容体を樹状突起上に輸送する役割があると考えられている。このGタンパク質非依存的な匂い応答は、多くの昆虫種で共通の匂い応答機構であると理解されている。昆虫の嗅覚受容体や味覚受容

体にはイオトロピックな機能が備わっていることや[030], [032], [038], ほとんどの動物で神経伝達物質の受容体として機能しているイオトロピック型グルタミン酸受容体が、昆虫では嗅覚受容体として機能していることが報告されるなど[036], 末梢における化学感覚の分子基盤が昆虫独特であることが次々と明らかにされた。

この受容体ファミリーは、旧口動物で出現しており、脊椎動物には保存されていないイオトロピック受容体は餌に含まれる様々な匂い物質に加え、酸やアミンを受容する[039], [040], [036]昆虫の嗅覚受容体は7回膜貫通構造を備えるが、N末端を細胞内に有する高次構造となっており、昆虫の嗅覚受容体はGPCRとは明らかに異なっている。

本研究に関するショウジョウバエの嗅覚システムは24種類の嗅覚受容神経細胞(ORN)を持つとされていたが、現在その時には同定されなかったイオトロピック受容体の存在も明らかになっている。すべてのイオトロピック受容体は嗅覚受容体ではないが、ショウジョウバエの嗅覚受容体はイオトロピック受容体であり、一連のリガンド応答型のイオンチャンネルを形成することが同定された。本研究ではこのショウジョウバエの嗅覚受容体について取得されたデータを解析対象とした。

### 1.8. 匂いセンシング空間の定義

空気中に存在する匂い物質は多くの場合混合物であり、ガスセンサ(QCM = Quartz Crystal Microbalance、無機半導体、有機半導体センサ)、分析機器(GC、LC(GC-MS 含む)、MS)、種々の動物における嗅覚受容体の包接挙動、嗅覚受容細胞群の神経応答、嗅球の活性化、脳の刺激による活性化、ヒトによる官能評価によって、数値化されたり、数値表現可能な物理現象を引き起こしたり、言語的に描述されたりする。

センサおよび分析器による手法、生理学的・分子生物学的手法を用いること、あるいは官能評価により匂いを多次元情報化できる。このようにして得られる情報空間を匂いセンシング空間と呼ぶことにする。

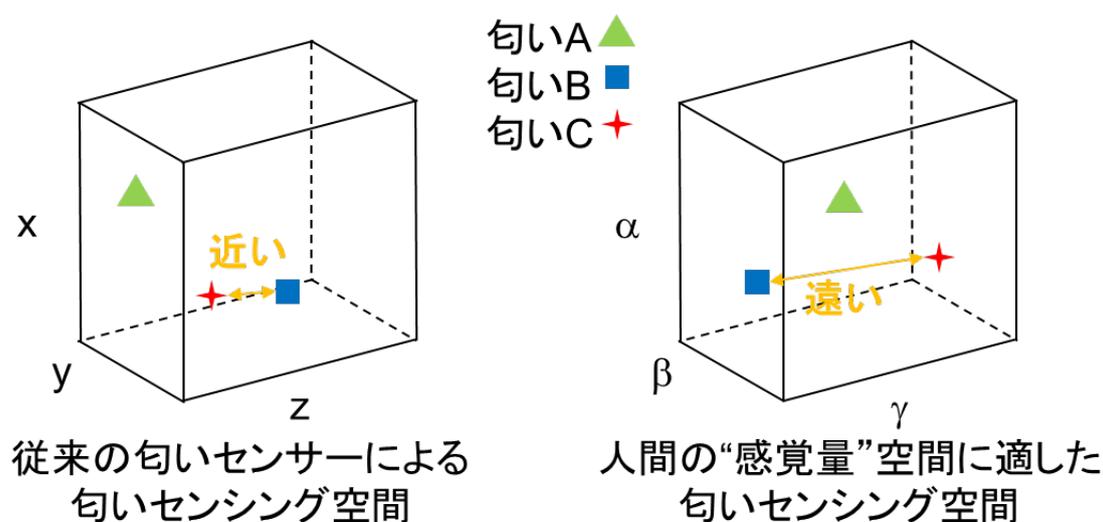


図 1-13. “感覚量”の空間

匂いを構成する香気物質の中には似たような印象のものも多い反面、微量で香気的印象に大きな影響を与える低閾値の香気物質も存在することが知られている。一部の化合物については、人間の感覚器によってガスクロマトグラフよりも鋭敏に検出される反面、分子構造が異なる化合物でも似た匂いとして感じることもあるこのように、生体の嗅覚システムは特定の物質群の差に対しては敏感だが別の物質群の差には鈍感だったりする。したがって揮発性の主に有機化合物の混合物である匂いの差異の評価において、その位置関係は図 1-13 に示すように異なると考えられる。分析装置により得られる匂いセンシング空間が xyz 軸からなる情報空間を持っているとし、人が感じているような匂いセンシング空間が  $\alpha\beta\gamma$  軸からなる情報空間を持っているとする。人が感じているような匂いセンシング空間の中で匂い B,C の距離は遠いが、それに対して分析装置により得られる匂いセンシング空間における匂い B,C の距離は必ずしも同じ距離になるわけではない。

本研究で筆者は、匂いを構成する香気物質の混合比率に基づく匂いセンシング空間は、匂いの印象を表現するには冗長なデータであると考え、“感覚量”に近い匂いセンシング空間による解釈が必要である。本研究では、特に生体の嗅覚受容の空間(“感覚量”空間)の解析をもとに、匂いの再現を考える。

## 1.9. 匂いセンシングデータの取得方法

### 1.9.1. 感応膜や多孔質などによる表面修飾したガスセンサ(QCM、無機半導体、有機半導体センサ)

これまでに複数の動作原理によるガスセンサ・匂いセンサが提案されてきた。この中でデバイス設計において何らかのチューニングをすることで、ある特定の測定対象のみに反応する選択的なセンサを作成できるものも提案されてきた。以下にその例の一部を示す。

半導体ガスセンサは還元性ガスの検出に利用されるセンサである。酸化スズ粒子を空気中に曝すと、酸素が粒子表面に吸着する。ガス(還元性ガス)中では、酸化スズ表面の酸素が還元ガスと反応して取り去られて、酸化スズ中の空乏層が失われ、電気が流れやすくなる。半導体ガスセンサはこの現象を利用しており、酸化スズ微粒子を利用しかつ数百度の温度まで加熱して使用するものが一般的である。

FET センサは、半導体素子である電界効果トランジスタのゲートの絶縁膜表面に吸着した検出対象物質の電荷を検出するデバイスである。ゲート絶縁膜表面の分子修飾(例えば抗体や糖鎖の固定化)によって、さまざまな物質の認識・検出を行うことができる。

QCM ( Quartz Crystal Microbalance) は、表面に物質が付着するとその質量に応じて水晶振動子の共振周波数が変動する(下がる) 性質をもち、ガスセンサ・匂いセンサとして様々な用途に用いられる。水晶振動子の電極表面に測定対象を吸着されるための層を製膜すると、ある特定の物質のみに反応する選択的なセンサを作成することができる。

以上のような感応膜や多孔質などによる表面修飾したガスセンサを組み合わせることで匂いを検出させることで、複数の人工的な匂いセンサによる匂いセンシング空間を得ることができる。このような匂いセンシング空間をもとにした匂い判別・識別システムには期待が寄せられている。

### 1.9.2. GC、LC(GC-MS 含む)

ガスクロマトグラフィー (Gas Chromatography, GC) はクロマトグラフィーの一種であり、気化しやすい化合物の同定・定量に用いられる機器分析の手法である。クロマトグラフィー (Chromatography) とは物質の大きさ・吸着力・電荷・質量・疎水性などの違いを利用して物質を分離・精製する技法のことを言う。ガスクロマトグラフィーのキャリアは不活性ガスが用いられ、液体クロマトグラフィーでは反応性を持たない有機溶媒や水、あるいはそれらの混合溶媒が用いられる。クロマトグラフィーを適用した結果は各成分ピークの移動速度として現れ、クロマトグラフィーにおいてカラム末端から時系列に現れる物質量をチャートに表したものをクロマトグラム (Chromatogram) と呼ぶ。

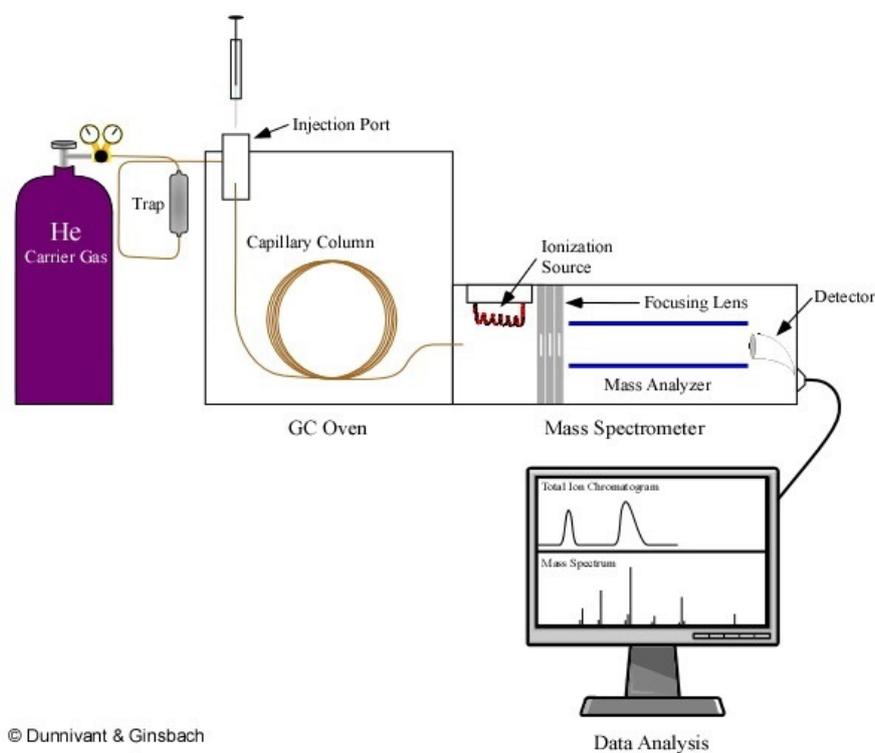


図 1-14. ガスクロマトグラフィーの基本構造 [042]

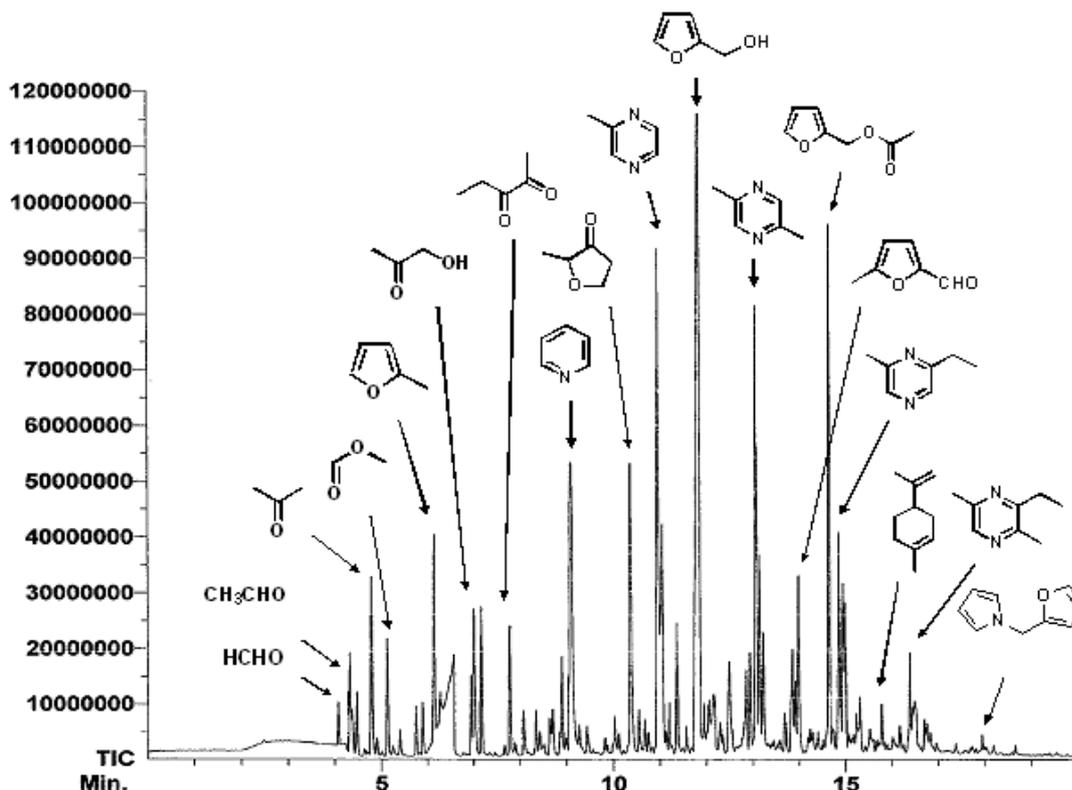


図 1-15. 香氣成分の分析例 [041]

市販のガスクロマトグラフィーの基本構造を図 1-14 に示す。試料はインジェクションポートから導入され、一般的なガスクロマトグラフィーにおけるキャリアガスとしては He がもちいられる。試料はキャリアガスにより長いカラムを通過するうちに分離し、成分により異なるタイミングでカラム終点に達する。カラムは正確に温調できるよう GC オーブン内におかれる。図 1-14 に示したガスクロマトグラフィーでは質量分析型の検出器が接続されている。一般に時間軸に対する検出器応答をプロットしたクロマトグラムを用いて全成分分析を行う。ピークの出現時間は保持時間と呼ばれる。

保持時間 (Retention Time) とは、サンプル注入時点から、分離された成分がピークを示す時点までの時間をいう。保持時間は分析条件が等しいかぎり物質によって固有の値である。ピーク面積 (Peak Area) とは、各成分ピークとベースラインとの間の面積のことである。たいていの検出器では、ピーク面積は成分濃度にほぼ比例する。

図 1-15 に、コーヒー香氣成分のクロマトグラムを示す[041]。モカコーヒーの香氣成分を室温にて、15 分間、SPME 法(固相マイクロ抽出:Solid Phase Micro Extraction)により抽出した時の GC-MS TIC(質量分析トータルイオンクロマトグラム;後述)を示した。分析の結果、アセトアルデヒド、アセトン、メチルフラン、ピリジン、ピラジン類、ヒドロキシメチルフラン等多数の微量化合物が検出された。コーヒーの香りは多数の成分による複雑な調和によって形成され、その数はおよそ 300 種以上とも言われている。その中でもコーヒー独特の強い香氣である焙焼臭のほとんどは、高感度な検出が困難な含窒素化合物、および含硫黄化合物によって構成されていることが特徴といわれる。

#### 1.9.3.0. 検出器

検出器はカラム出口に設置され、サンプルの各成分を検知して電気信号に変換する部位である。カラム末端で、物質量の出現確認に用いられる検出器にはいくつかの種類があり、汎用目的には TCD もしくは FID が用いられる。以下に説明する。

##### 1.9.3.1. FID(Flame Ionization Detector, 水素炎イオン化型検出器)

物質を水素炎中で燃焼することによって発生するプラズマ電子を検知するものである。C-H 結合(ただしカルボニル炭素と直接結合した水素は除く)を持つ化合物に対して感度を有するため、一般の有機物に対する感度は高いが、水や二酸化炭素などの小分子ガスは感知できないのが欠点である。基本的には破壊的な検出方法であるため、サンプルの分取は不可能である。

##### 1.9.3.2. TCD(Thermal Conductivity Detector, 熱伝導度型検出器)

物質の熱伝導度の違いを利用してサンプルの検出を行う。キャリアガス以外のほぼあらゆる物質を検知できるが、感度があまり高くないので希薄サンプルには不向きである。基本的には非破壊的な検出方法であるため、サンプルの分取が可能である。

##### 1.9.3.3. MSD (質量分析型検出器)

質量分析も感度の高い検出法である。質量分析検出を用いると、フラグメントパターンの類似性を用いて候補化合物を求めることができる。[043]

#### 1.9.4. 官能評価を併用したガスクロマトグラフィー

「匂い嗅ぎ GC (Sniff-GC) は、キャピラリカラムの出口を分岐し、一方を検出器に、他方を匂い嗅ぎポートに接続し、検出器によるクロマトグラムの取り込みと各ピークの匂い嗅ぎ(人間が嗅ぐ)を同時に行い、各成分の匂いの質を明らかにする」ための装置である。匂い嗅ぎ GC (Sniff-GC) はガスクロマトグラフィー—オルファクトメトリー(GC-オルファクトメトリー、GC-O)という呼称が使われることもある。検出器によるクロマトグラフィー結果が「クロマトグラム」と呼ばれるのに対して、匂い嗅ぎによるクロマトグラフィー結果を「アロマグラム」と呼ぶ。

AEDA(Aroma Extract Dilution Analysis)法とは、無臭空気で段階的に希釈し匂い嗅ぎGCを行い、最も希釈されたサンプルでもニオイとして検知された成分を探す方法である[044]。希釈率が大きくても匂いの強度が強い匂い物質を決定でき、匂いの強度(寄与率)を数値化することが出来る。

OASIS(Original Aroma Simultaneously Input to the Sniffing port method)および AROMATCH® [045], [046]は曾田香料が開発した、GC-O の出口から出てくる香気と原材料の香気を気相中で混合して評価する手法である。GC は香気成分がそれぞれ特定の時間で分離され出てくるので、それを分離され出てくるタイミングで、分析対象の香気と混ぜ合わせた時の香気のイメージが、分析対象とする香気と比べてどのように変化しているかを官能評価する手法である。特定の香気化合物、特にその中でも微量成分の分析対象とする香気における役割の解釈に役立つ。

### 1.9.5. MS(質量分析)

質量分析においては、試料はイオン化され、電界または磁界中で分離される。また、試料分子をイオン化する過程は要するに分子からの電子移動であるため、後続反応によって試料が分解することがある。この過程をフラグメンテーションという。フラグメントイオンの生成パターンは分子の構造によって(おおまかにではあるが)決まっており、分類・理論化が行われている。

マススペクトル (Mass Spectrum, MS) は、質量分析の結果得られる、横軸に質量(正しくは  $m/z$  値)、縦軸に検出強度をとったスペクトルである。試料分子の構造に関する情報が多く含まれるため、既知物質の同定や新規物質の構造決定に用いられる。二章の研究では MS を使用した。[047]

### 1.9.6. 官能、生理学的・分子生物学的手法

#### 1.9.6.1. Dravnieks の多様な香気物質に対する官能評価データ

Andrew Dravnieks による「ATLAS of ODOR CHARACTER PROFILES」[048]は、ある化学物質のにおいを嗅いだとき、どの様な記述子が適切かを調査した結果がまとまっている。160 種類の化学物質について、約 50 名のパネルからなる官能試験を行い、146 の記述子(Fruity や Spicy などのキーワード)について 6 段階評価(0~5 点)し、どの様に記述するパネルが多いかが掲載されている。「ATLAS of ODOR CHARACTER PROFILES」の評価パネル群は、匂いトレーニングを受けており、提示された匂いに対して官能評価用語を十分使いこなせる評価パネル群である。Dravnieks の Database は、人の匂い認識をデータとして大規模に収集したデータセットを提供してくれる。解析することで匂いの知覚上の特性を大規模に調査することができ、官能評価による「匂いセンシング空間」と呼べる。

#### 1.9.6.2. fMRI による脳の匂い応答

MRI (Magnetic Resonance Image) とは、核磁気共鳴現象 (nuclear magnetic resonance: NMR) という物理現象を応用して、人体の断面などを画像化する手法である。MR 信号(核磁気共鳴現象によって発生する信号)を透過させ、Image 化(画像化)できるため、外科的な手術を行わずに私達の体や脳の断面像を得ることができる非侵襲的手法である。

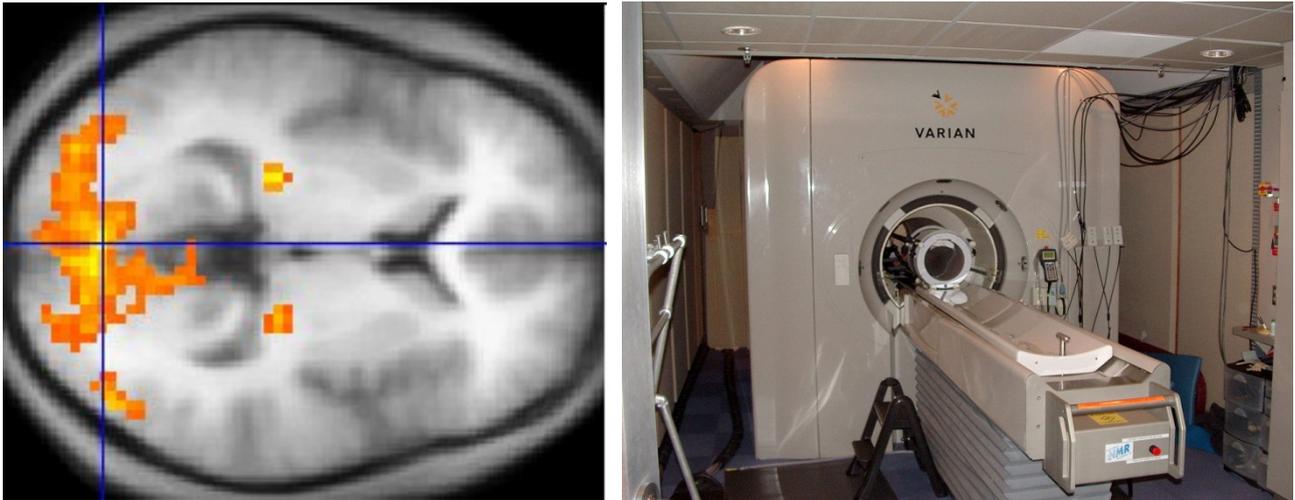


図 1-16. MRI で取得できる断面図と、装置外観 [049]

外部からの刺激や課題を行うことによって活動した脳など臓器の様子を、この MRI を使い画像化する方法を特に fMRI とよぶ。装置と得られた観測結果の解析例を図 1-16 に示す。脳活動に伴って「脳血流量」や「脳酸素代謝率」が変化することで、MR 信号が変化する。神経細胞が活動すれば、それだけ多くの血液が流れ、酸素を消費するので、血流量が増え、酸素を失った還元型ヘモグロビンも増える。血液中で酸素を運ぶヘモグロビンの、酸化型と還元型のバランスの割合である「脳酸素代謝率」を神経細胞の活動として、MRI を用いて脳の活動（機能）を画像化するこのように、神経細胞の活動→脳血流量・脳酸素代謝率の変化→MR 信号の変化、というように脳活動を間接的に反映したものが、fMRI で画像化する「MR 信号の変化」である。

神経細胞の大きさは1ミリ以下、そしてその一回の活動は数ミリ秒で起きるとされている。脳の活動を測定するどの方法にもつきまとう「時間分解能」と「空間分解能」の問題に対して、fMRI の時間分解能は数秒、空間分解能は数ミリから数センチといわれている。fMRI では、神経細胞レベルの活動まで測定することは、いまのところ非常に難しいようである。

Jean-Pierre Royet らは f-MRI を用いることで、匂いに関連した心の中でのイメージが、匂いに関する脳部位の神経基盤へのインパクトを見積もった。匂いの機能に関する脳の領域へのその専門職歴の効果を見積もるために、初心者と経験のある調香師が、匂いを知覚するとき・想起するときに、どのように脳が活性化するか調べた[049]。すべての被験者で匂いのイメージが第一嗅覚皮質(梨状皮質)を活性化させ、匂いの知覚・想起において近い神経器質が活性化されることが分かった。だが、嗅覚の訓練が前部梨状皮質、眼窩前頭皮質、海馬に影響を与えていることもわかった。;生徒のグループでは、意識的な知覚をつかさどる前頭葉に活動が集中していたが、ベテラングループでは、記憶の想起や心的イメージをつかさどる海馬傍回でニューロンの活発な動きがみられた。調香師の専門知識・技術の習得に従い、嗅覚を機能させる脳内活動が、前部梨状皮質、眼窩前頭皮質、海馬(呼吸や嚥下など自動的身体機能をつかさどる領域)に、徐々に移動していくことも明らかになった。調香師の専門知識・技術の習得過程で、嗅覚と記憶の脳領域で機能の再構築が起こり、その再構築が匂いの想起やフレグランスの創作の類いまれな能力を裏付けていると考察した。[049]

### 1.9.6.3. マウスの嗅球マップ

昆虫およびほ乳類における匂いが脳内にエンコードされるかを理解することは、神経生物学の基礎的重要事項である。嗅覚情報処理の最初の 2 つのステージは、脊椎動物および無脊椎動物の両方で、比較的良好に研究されている。しかし、高次嗅覚表現の組織的な原則は、よくわかっていないままである。

マウス嗅覚系では、匂い物質が嗅覚受容体神経 (Olfactory Receptor Neuron:ORN、嗅神経細胞:olfactory sensory neuron: OSN とする記述方法もあるが、ここでは特に区別せず ORN に統一する)に存在する数千種類の異なる嗅覚受容体(Olfactory Receptors:ORs)によって検出される。各 ORN は唯一の機能する OR 種を発現し、「1 ニューロン 1 受容体」ルールと呼ばれている。さらに、嗅球(OB)の特定の投影サイトである糸球体に、同一の OR を持つ OSN 軸索は収束する。すなわち、「1 糸球体-1 受容体」ルールをもつ。これらの基本的なルールに基づいて、OSNs によって検出された匂い物質の結合シグナルは嗅球(Olfactory Bulb : OB)の中の糸球体マップ、活性化された糸球体の「地形情報」に変換される。マウス嗅覚系における神経地図形成は新生児期の活動依存的に依存していることも報告されている。[050], [051]

生体内の活性部位を可視化するイメージング法は近年進展を見せている。たとえば特定の糸球体活性化部位を可視化したトランスジェニックマウスを用いることで、大まかな匂いの種類の境界と糸球体活性化部位の境界をマッピングできる。図 1-17 に示したのは、benzaldehyde / citronellol / ethyl butyrate によりマウスの活性化される嗅球の背側表面上の糸球体をイメージング法で解析した結果の一例である。図中において赤や黄色で示される糸球体部位が強く活性化していることを示す。大まかな匂いの種類と糸球体活性化部位との対応関係が個体間で一致することを見出され、また、マウスの匂い地図がラットの匂い地図と種を超えて同じような配置になっていることも見出されている。‘大まかな匂いの種類’に対して活性化する部位が遺伝的に規定されている可能性、あるいはその程度について議論されている。

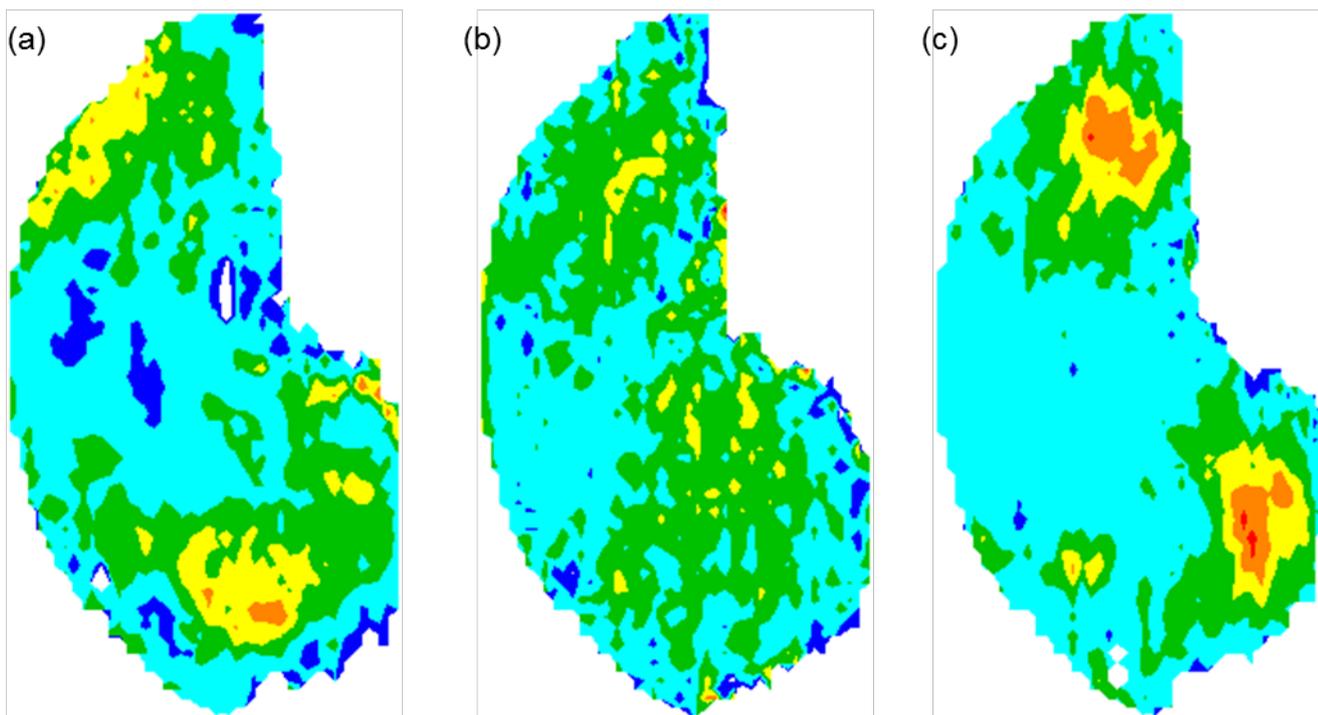


図 1-17 マウス嗅球の糸球体のイメージング法による活性化測定例 ; (a) benzaldehyde (213 ppm in air) (b) citronellol (2.5 ppm in air) (c) ethyl butyrate (75 ppm in air) [052], [053]

#### 1.9.6.4. 昆虫の嗅球マップ

嗅覚系の最初の情報伝達経路で神経細胞は各嗅覚受容体に対応する糸球体に分離する。嗅覚受容体は、脳内で、糸球体において正確に投射される(「チャンネルコーディング」)。ショウジョウバエキロショウジョウバエ及び他の昆虫で、ラテラルホーンとキノコ体の上位嗅覚脳の 2 つの領域が存在する。これらの領域は、匂い物質をエンコードする機能と彼らが使用するメカニズムに特化していると考えられている。キノコ体は、嗅覚学習能力の基礎となる一方、ラテラルホーンがそれは未学習の匂いが直接結びつく行動のベースとなるような計算を実行することが示唆されている。キノコ体において匂い物質への反応をラテラルホーンニューロン (LHNs) 糸球体間でどのように情報を統合する機構についても明らかになりつつある。[054]

#### 1.9.7. 嗅覚受容体を用いた匂いセンシングデータの取得方法

##### 1.9.7.1. 嗅覚受容体の分子生物学と受容特性

嗅覚は、哺乳類の生存に不可欠である。環境における多様な匂い分子は、鼻腔の嗅上皮に発現する嗅覚受容体(ORs)によって検出される。ホストゲスト科学の観点から、嗅覚受容細胞のブロードな応答について考えると、ゲストの包摂挙動において、複数の水素結合のような認識部位が必要となる。加えてエステルや脂肪族化合物はゲスト分子として構造が比較的柔軟なので、そもそも高い選択性が得られにくい化学刺激物質である。実際、嗅覚受容細胞の挙動を観測した結果、フェロモンのような特異的な応答ではなく、ブロードな応答を示すことが分かっている。

OR 遺伝子の数が彼らの生活環境に応じて、それぞれの種における偽遺伝子が高い割合 (20-60%) 存在し、OR

遺伝子の個数が、動的に、頻繁な遺伝子重複と擬遺伝子化イベントを伴い、進化の過程で変更されていると考えられ、([054]- [058])生物間で高度に可変的であるものの、一般に、哺乳類のゲノムは、～1000 個の OR 遺伝子を保有し、哺乳動物における最大の多重遺伝子ファミリーを形成する。

嗅覚細胞群がとある香気物質を受容したとき、どのような応答群を返すのか、つまり受容体の香気受容や脳内での神経信号の変換の様子、香気刺激の神経応答空間内マッピングを「コーディング」と表現できる。生物種ごと、各受容器、それぞれの場合において、香気受容の匂いセンシング空間は当然異なると考えられる。

### 1.9.7.2. 電気生理学的手法

電気生理学 (electrophysiology) とは神経細胞などの興奮現象を電氣的に検出する生理学の手法である。ガラス管電極や金属電極を生きた神経細胞内に刺入し、微小な電流や電位はアンプ (amplifier) で増幅し、オシロスコープなどの計測機器で細胞の電気現象 (電流・電位) を記録する方法である。記録したデータはコンピューターで解析する。

1980 年代から遺伝子クローニング 技術と電気生理学が合わさり、イオンチャネル遺伝子を別の細胞に強制的に発現させて、その働きを調べる方法が確立した。そのためによく用いられる方法のひとつがアフリカツメガエル卵母細胞系を利用したものである。パッチクランプ法も電気生理学的方法の一種である。その典型例の手順を図 1-18 に示す。細胞膜に発現している受容体の機能を調査することができる。技術的には非常に難しく、熟練の技が必要となる。

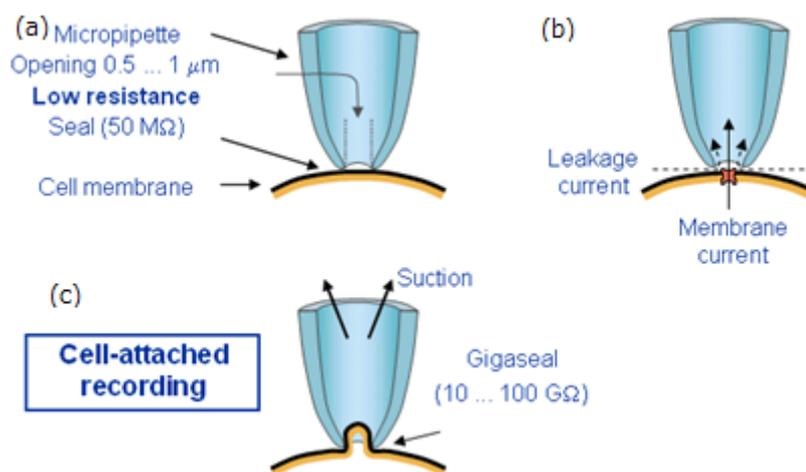


図 1-18. 典型的なパッチクランプ法の例 ; (a)マイクロピペットを細胞に近づけた状態、(b)漏れ電流の発生、(c)ギガシール形成

[059]

1. ガラス電極に実験目的に応じて選択された電極内液を満たし、金属線が付いた電極ホルダーへ電極を装着する。電極内液と接触している金属線はプリアンプへ接続する。

2. 電極先端の抵抗を矩形波電流をモニターしながら、細胞へのアプローチと接触を行う(図 1-18(a))。
3. 顕微鏡とマニピレーターを操作し、標的細胞へ電極先端を近づける。細胞に電極先端が近づくと矩形波電流が小さくなる。また、陽圧によって細胞表面にくぼみ(dimple)が観察される。
4. 細胞膜が電極先端へ接着したときに吸引に切り替え(図 1-18(b))、細胞膜とピペット先端を密着させ、ギガシール( $1\text{G}\Omega$ 以上の抵抗)を形成する(図 1-18(c))。このギガシールにより、漏れ電流(leak current)が非常に小さくなり、イオンチャネルの開閉による小さな電流を測定することが可能となる。

#### 1.9.7.3. Hallem ら調査—香気物質への応答挙動

ショウジョウバエの嗅覚系においてはHallemらが24種類の嗅細胞の応答データを報告した[060]。自然界に一般的に見いだされる110種類の香気物質に対するショウジョウバエの24種類のORNの興奮性・抑制性応答が取得された(3章参照)。プラスは興奮性の応答(数が多いほど強い興奮性)を表し、マイナスは抑制性応答を表す。

#### 1.9.7.4. 指示薬利用法(染色法)

蛍光色素を用いた細胞内イメージング：細胞内に $\text{Ca}^{2+}$ などを感じ取る各種の蛍光プローブを導入し、その濃度変化を2次的に観察する方法である。この方法では、細胞内で起こっていることをリアルタイムで画像として追っていくことができる。

#### 1.9.7.5. 神崎研究室セルベースバイオセンサー

現在神崎研究室において図 1-19 に示すように昆虫のイオノトロピックレセプター遺伝子を $\text{Ca}^{2+}$ 感受性蛍光タンパク質遺伝子と共にSf21細胞へ導入した、バイオセンサーとして利用可能なヘテロ遺伝子発現細胞の作製が報告されている[061]。同研究室では、図 1-19(A)に示すような手法でセンサ細胞を作成し、カビ臭の中に含まれる1-octen-3-olに最も強く応答するOR13aの細胞株種の香気への応答を取得した。Sf21母細胞へヘテロ遺伝子発現させるための2種類の発現ベクターをループで模式的に表している。片方にはOrcoおよびOR(この場合はOR13a)、もう片方にはGCaMP3を遺伝子挿入するための発現ベクターを構築する。母細胞へ導入し、両タンパク質が発現する安定株種を作成する。これに成功すると図 1-19(B)に示すような細胞を得られる。1-octen-3-olを嗅覚受容体が受容すると、イオンチャネルが開き、カルシウムイオンが細胞内に流入する。細胞内にはカルシウムイオンに配位することでその蛍光強度が強くなる、 $\text{Ca}^{2+}$ 感受性タンパク質GCaMP3が存在しており、細胞の蛍光強度が増加する。蛍光強度を追跡することにより細胞を匂いセンサとして利用することができる。

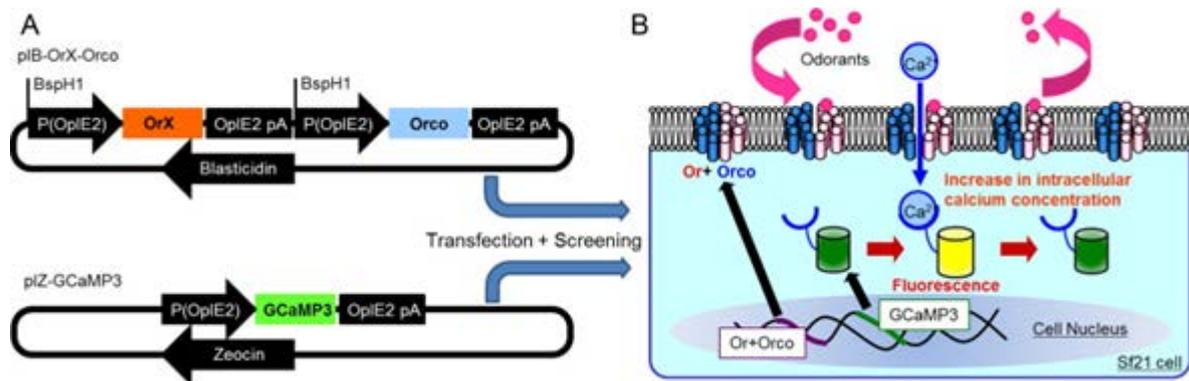


図 1-19.昆虫のレセプター遺伝子を導入した、バイオセンサー細胞の作製；(A) Sf21 母細胞へヘテロ遺伝子発現させる発現ベクター、(B) 安定株種細胞の動作 [061]

流路を用いて匂い物質濃度変化に応じた蛍光変化を追跡することにより、ORN 応答挙動の取得が行われた。嗅覚受容体の匂いセンシング空間のデータ収集において、自動的に試薬の分注と励起光照射による蛍光測定が可能となる蛍光プレートリーダーによって、嗅覚受容体の応答データの収集が大規模に可能となると考えられる。

#### 1.9.7.6. 蛍光プレートリーダー(FDSS)を用いた昆虫の嗅覚受容体応答測定について

マイクロプレートリーダー (Microplate Reader) とは物理学・化学・生物学の実験や検査などで広く用いられる測定用機器で、マイクロプレートに入れた多数のサンプル (主として液体) の光学的性質を測定するものである。典型的な内部の構造を図 1-20 に示す。分光測定の対象としては最も一般的な吸光のほか、蛍光、化学発光、さらに蛍光偏光などがあり、モードが切り替えられる装置もある。特に現在では測定温度を一定に保ち、連続的に時間変化を測定できるものも多い。

生物学的なスクリーニングなどの分野 (創薬開発や、基礎研究分野での生細胞の活性研究) における活用が期待されている。異なる種類や濃度の試薬を全ウェルに同時に分注し、蛍光や発光の時間変化を全ウェル同時に測定できるためである。ウェル間の時間差無く、短時間で大量に測定することが可能である。実際、光学的測定方法はハイスループット性が高く、特に近年では発色プローブや蛍光プローブを用いて測定する方法が発展したため、かつて放射標識を用いていた試験に代わり、様々な実験・検査に必要なものとなっている。本研究において、第五章の一部と第六章で用いたデータは蛍光プレートリーダーを利用して得られたものである。

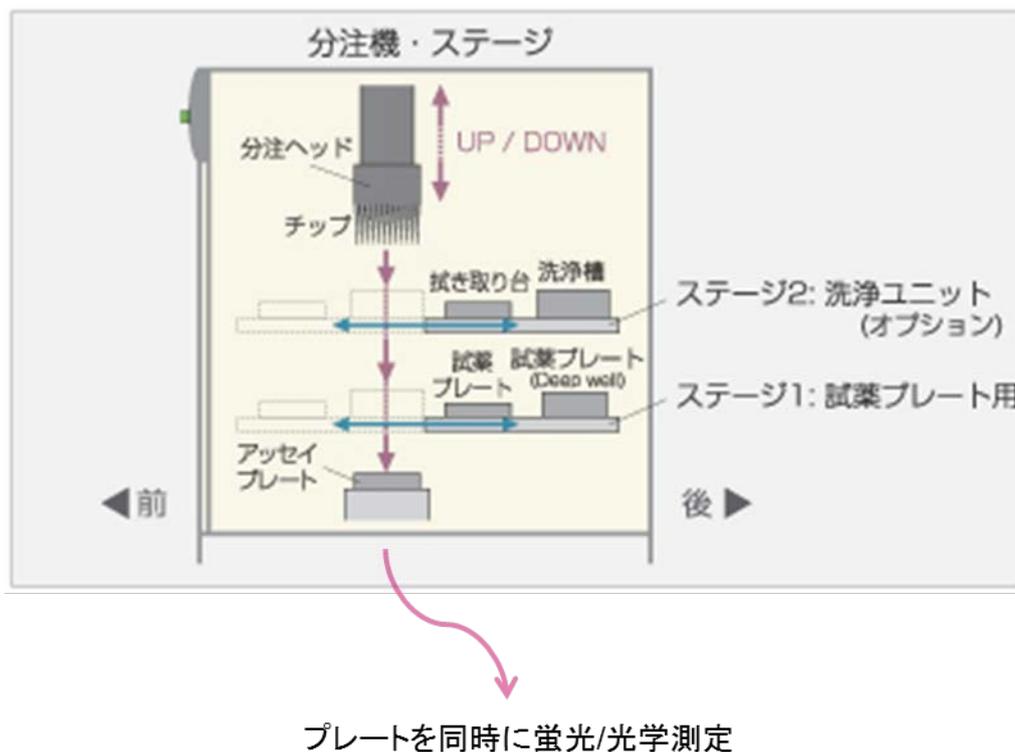


図 1-20. 蛍光プレートリーダー(FDSS)を用いた昆虫の嗅覚受容体応答測定[062], [063]

#### 1.9.7.7. ヒト嗅覚受容体による匂いセンシングデータ取得

分子生物学や薬学において、細胞内メッセンジャーの働きを強める、生体の作用を強めることによって作用する薬を「アゴニスト(作動薬)」という。また、メッセンジャーの働きを阻害する場合、生体の作用を弱めることになる。この作用によって作用を発揮する薬を「アンタゴニスト(拮抗薬)」という。つまり、アゴニスト(作動薬)は活性化させる物質の総称であり、アンタゴニスト(拮抗薬)は不活性化させる物質の総称である。

嗅覚受容体において受容されるアゴニスト・アンタゴニストが明らかになっていない嗅覚受容体はオープンな受容体、嗅覚受容体のアゴニスト・アンタゴニストの決定は脱オープン化と呼ばれる。

人間の嗅覚系は膨大な数の匂いを判別することができるが、我々は現在、どのような化学的特徴が、嗅覚受容体によってコードされているのかを理解していない。ヒトは 400 種類弱の欠損しておらず発現をする嗅覚受容体を持っており、個々の遺伝的変異した固有のセットとなっていて、嗅覚受容に多様性を導いている。膨大な種類の匂い分子と何百もの受容体との相互作用をカバーしようとする、探索空間内におけるデータの不足が問題となる。

デューク大学の松波らは、異種ルシフェラーゼアッセイを使用した嗅覚受容体のアゴニスト・アンタゴニストの特定を進めているが、511 種類の嗅覚受容体のクローンライブラリーに対して、73 種類の匂い物質をスクリーニングしたのみであり、リガンド解明されたものは唯一の 10%にとどまっている。[021]

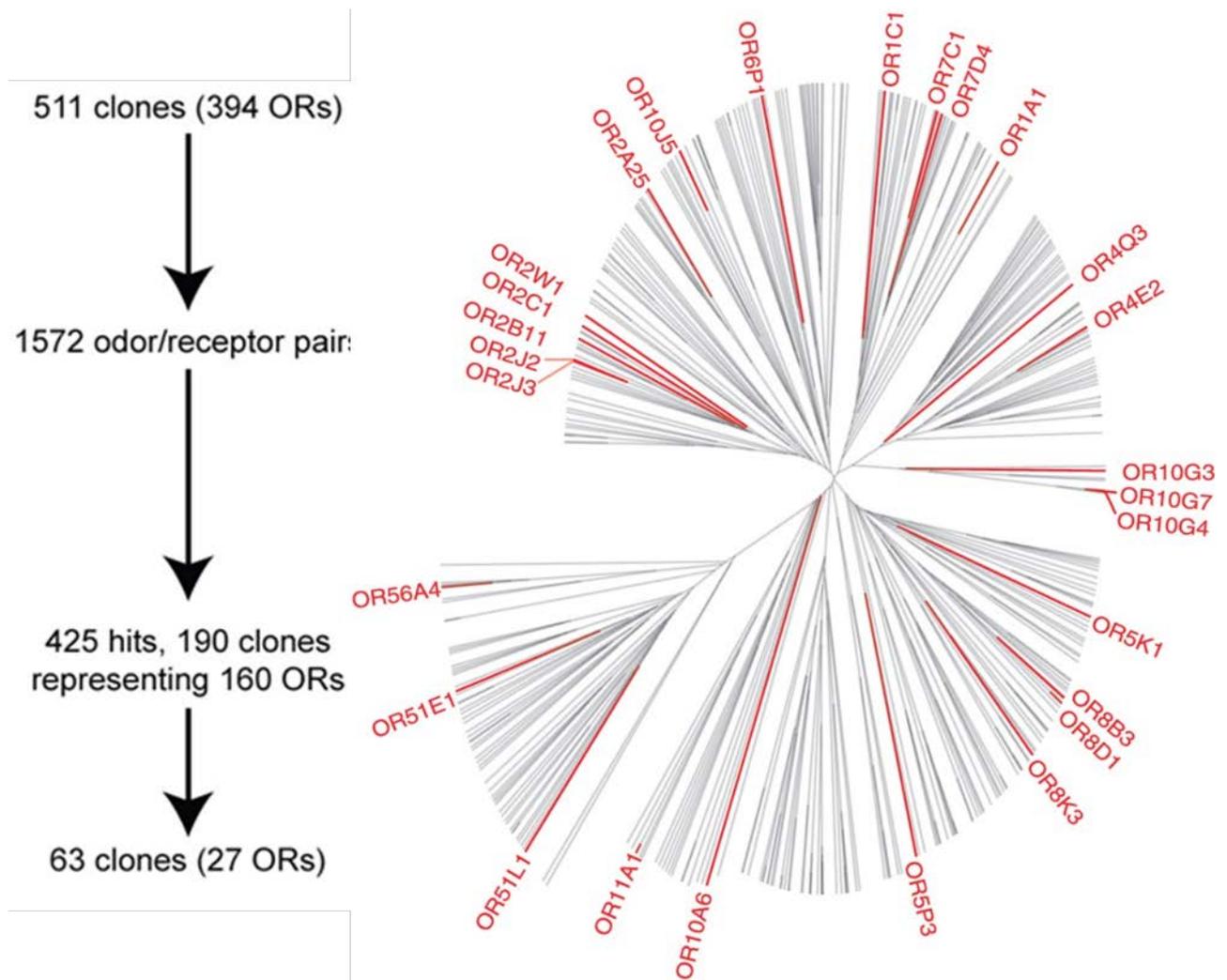


図 1-21. 27 種類の嗅覚受容体のクローンライブラリーに対する 73 種類の匂い物質をスクリーニング、右の樹状図は遺伝子の相動性に基づく各受容体の類似性を可視化したもの。赤で示す受容体名が脱オーファン化された受容体 [021]

これにより、図 1-21 に赤で示す受容体名が脱オーファン化された 27 種類の嗅覚受容体のクローンライブラリーに対する匂い物質が可能となっている。遺伝子の相動性に基づき各受容体の類似性を樹状図にて確認すると、脱オーファン化されている ORN の種類に大きな傾りはないと思われる。だが、ほ乳類の嗅覚受容体のデータは限られており、本研究で利用するデータとしては不十分であると考え、使用しなかった。

## 1.10. 多次元データ解析を用いた解析

### 1.10.1. Dravnieks の官能評価データに基づく検討

#### 1.10.1.1. Noam Sobel らの嗅覚におけるホワイトノイズの定義

視覚では、二種類の(刺激)混合物、独立した多くの異なる波長を含む各セットは、「白」という共通の色知覚を生成することができる。聴覚では、二種類の(刺激)混合物、独立した多くの異なる周波数を含む各セットは、「ホワイトノイズ」と呼ばれる一般的な知覚におけるハムのような状態を生成することができる。視覚と聴覚の「白」は、混合成分が、刺激領域にまたがる場合と、その成分が等しい強度の場合の二つの条件が当てはまるときにみられる。

香気物質の混合物にこれらの同じ条件を適用した場合、嗅覚でも同様に「白さ」が出現するのではないか、という仮説を、Noam Sobel らは立てた。[064]

Noam Sobel らは、嗅覚刺激空間の広がりによって 86 の分子を選択し、それぞれほぼ等しい強度になるようにそれらを希釈した。Noam Sobel らは、次にそれぞれ種類の分子成分を含む、様々な匂い物質の混合物を調製し、このような複数の混合物の知覚的類似性を、人間の参加者により評価したところ、匂い物質の混合物が等しい強度成分であれば、私たちは非重複な成分数を増加させたときに、共通の単一のコンポーネントを持っていない場合でも、混合物は、互いにより類似になることがわかった。約 30 種類のコンポーネントを使用すると、混合物はほとんどの似た匂いになる。参加者は、新規な任意の名前の付けられた等しい成分強度にした約 30 種類からなる混合物をよく嗅いだ後、より少ない成分数でできた混合物でも、等しい成分強度にした約 30 種類からなる刺激空間にまたがる混合物を同じタイプの匂いであると知覚した。

以上の実験から Noam Sobel らは、一般的な嗅覚知覚、嗅覚の「白」は、刺激空間にまたがる、約 30 種類以上の同強度成分の混合物と関連しており、特定の分子が特異的に嗅覚表現に寄与するというよりは、むしろ嗅覚表現は分子の特徴に影響されることが示唆されたと結論した。[064]

#### 1.10.1.2. Jason B. Castro らの匂い認識空間のクラスター化解析

Dravnieks のデータは、嗅覚受容の構造の洞察を得るために刺激の大規模なデータベースである。嗅覚認識を分析・マップした結果から、嗅覚知覚空間は次元が高い証拠が得られている。これらの結果に加えて、Jason B. Castro らは次元圧縮の方法として使用可能な数学的手法である NMF(Nonnegative Matrix Factorization: 非負値行列因子分解)法[080]を用いて、多次元表象空間から、評価パネルに置ける匂いプロファイルの構造を解析したカテゴリー化に使用する香りの次元について解析した。嗅覚知覚空間の次元の問題を検討する上で、新たな視点を使用し、次元の問題についてもより厳密に検討し、嗅覚の空間の次元の上限の推定を行った。[065]

Jason B. Castro らは、NMF 法を適用し、人間の嗅覚知覚について分類を試みた。匂い物質を円滑に対離散的に、描述子空間への配置を調査するために、NMF 法を適用し得られた各要素行列に対応する記述子もしくは香気の関係性を調査した。SNE(stochastic neighbor embedding)アルゴリズムを使用して、基底行列  $W$  と係数行列  $H$  に対して SNE を構築した。これにより、例えば、元の空間に接近している匂いと一緒に近くに表現されるというような、相対的な相互距離関係は保持されたまま元の高次元空間における匂いの間の全ての距離の、平面表現が得られる。描述子空間 ( $W$ ) に SNE を適用することで、図 1-23(左)に示されている 146 記述子の、8 つの離散的で非重複なクラスターが得られ、同様に、匂い物質のスペース ( $H$ ) に SNE を適用し、図 1-23(右)に示す 144 個の匂いの 10 個の離散的で非重複なクラスターが得られた。空間に対して NMF 由来の知覚描述子は連続的に占有されていないことが示された。

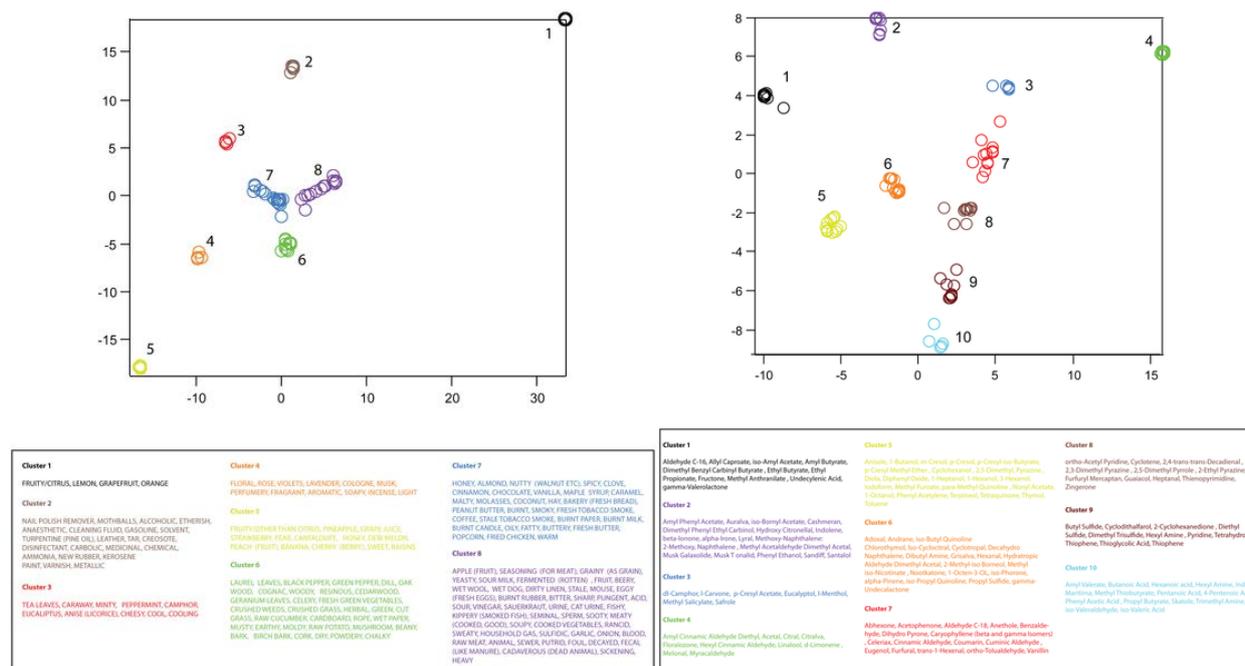


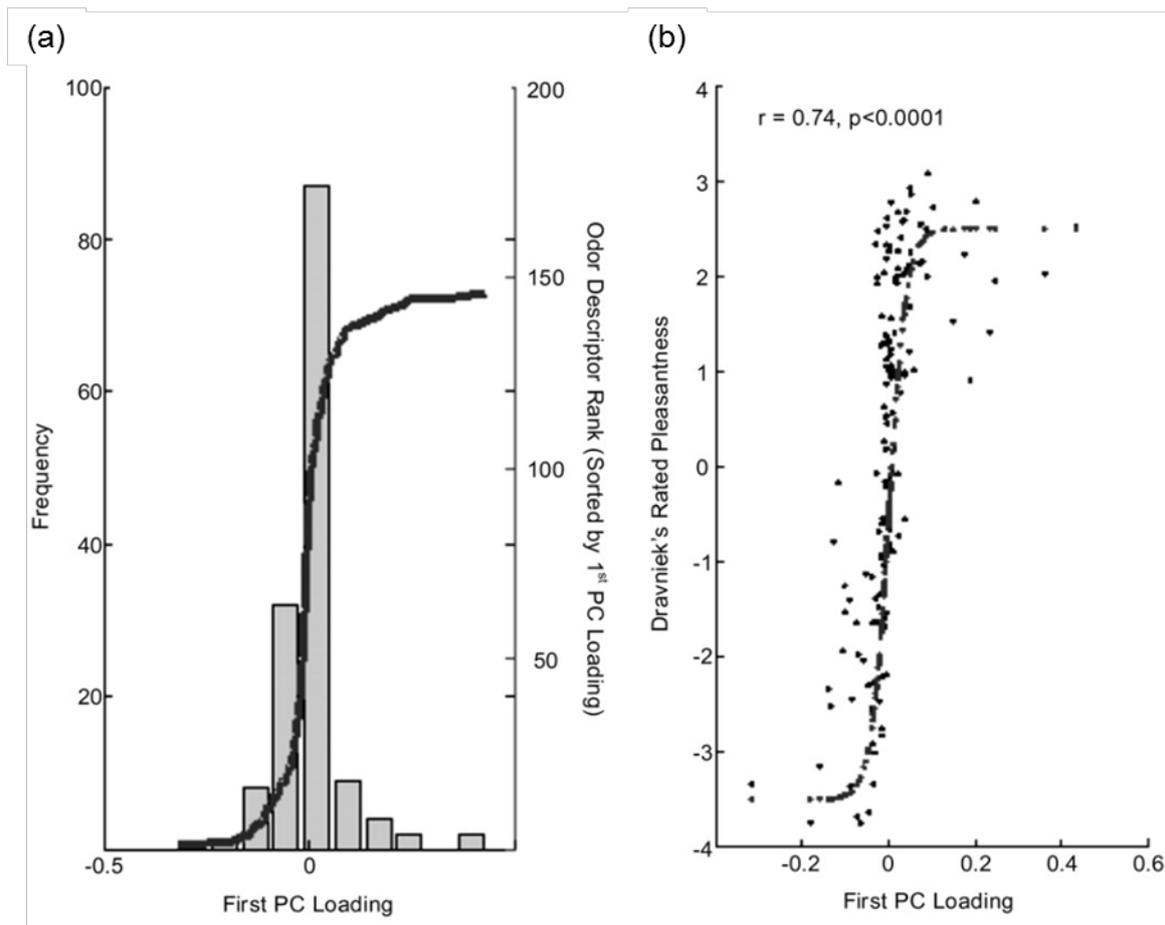
図 1-23. 確率隣接埋め込み(SNE)アルゴリズムを使用した描述子空間・化合物空間への配置決定、左; Two-dimensional embedding of the descriptor-space, 右; Two-dimensional embedding of the odorant-space, [065]

一般的には NMF による解析に対して、一意に決定されるベクトル組合せが存在するわけではない。部分空間の大きさの複数の正当な選択肢があるが、これらの各々からデータに関するさまざまな洞察が可能であり、または別の目標のために役立つ。

Jason B. Castro らは、誤差分析から、最適なカットオフ次元数として、10 次元が良いと考え、その論文内で個々の臭気プロファイルの分類、全ての抽出された基底の匂いセットに対する発現を評価した。人間の香りの受容は 10 種類くらいの香りとして分類され、それはある程度普遍性のある分類になっているようだ結論した。いうまでもなく、より大きなデータセットを使用することで香り受容の質的評価を化学構造から推測できるようになると考えられるのだが、このデータに対して NMF 法を適用し、匂い空間は連続的な知覚ではなく、クラスターに分けることができると考えた。

### 1.10.1.3. 匂いの認識空間の主成分分析と匂いへの嗜好性の関係

匂いは、比較的小分子量の香気物質で作られている[066]ことが分かっている。分子の物理化学的特徴が知覚上の匂いのイメージを決定することは明らかであるが、この関係を規定する規則は不明のままである。カーンらは理解を困難にする重大な障害は、知覚と分子の両方を情報空間が高次元であることであるとした。



**図 1-24. Dravnieks データの第一主成分と知覚の快不快の相関性の検証 ; (a)主成分分析の値によってソートした重み、(b)主成分分析のコンポーネントの重みの快不快に対するプロット [067], [068]**

前節でも用いられた Dravnieks のデータは 144 種類の匂い物質及び 146 種類の間の評価用語のラベルの行列から生成された知覚空間となっており、まずカーンらは主成分分析 (PCA) を用いて、少数の主成分 (PCs) を導出することで、この多次元空間の次元を低減した。

Dravnieks のデータは 144 種類の匂い物質及び 146 種類の間の評価用語のラベルの行列から生成された知覚空間となっており、まずカーンらは主成分分析 (PCA) を用いて、少数の主成分 (PCs) を導出することで、この多次元空間の次元を低減した。

知覚における第一主成分 PC1 が、匂いの快を反映しているかどうかのテストが行われた。カーンらは主成分分析上のコンポーネントの重みを使用して 146 知覚記述子のそれぞれについて、快評価を公表して比較している。図 1-24(a)に示したものは主成分分析の値によってソートした重みである。図 1-24(b)に示したのは主成分分析のコンポーネントの重みとの関係を快不快に対してプロットしたものである。カーンらはシグモイド変換を併用しながら回帰分析を行い、Dravnieks のデータベースにおける評価用語の主成分分析の第一主成分と快不快が高い相関を持っていることを確認した ( $R=0.74$ ;  $F(1,144)=178.4$ ;  $P<0.0001$ )。

図 1-24 に示した知覚における第一主成分 PC1 と匂いの快との相関性テストに加えて、カーンらは匂い刺激の空

間は、多量の物理化学的記述子を用いた情報空間としての解析も検討した。カーンらは多数の匂い物質は、Dravnieks が得たような口語的記述子による情報空間を形成していると同時に、PubChem のような化学物質データベースを用いて分子構造から物理化学的記述子へと変換・計算出来ると考えた。物理化学的性質をもとにして、新規な分子の快不快を予測することができたとしている。[067]

Nathalie Mandairon は匂いの快不快が匂い分子の構造により予め定められていると仮定して、マウスなど他の哺乳類も、人間に似た匂いの好みを示すと予想した。

Nathalie Mandairon らの匂い物質の選択は、カーンらによる最近の研究に基づいた。カーンらは主成分分析 (PCA) を用いて、二つの匂い物質情報空間を生成した: Dravnieks のデータは 144 種類の匂い物質及び 146 種類の間の評価用語のラベルの行列から生成された知覚空間となっており、PCA を用いて少数の主成分 (PCs) を導出することで、この多次元空間を低減できる。各 PC (主成分) は、連続する各 PC が最大可能分散となるよう順序付けられている。カーンらの Dravnieks のデータへ主成分分析 (PCA) を適用した匂い物質情報空間を元にした匂い分子の構造により予め定められている快不快それぞれの香気物質を用いて、それが実際にマウスの行動研究でも、快不快評価は匂い分子の構造と関連していたことを Nathalie Mandairon らは示した。

嗅覚嗜好が実際に一部の匂い分子の構造に起因していることが示されたと結論付けており、この研究はカーンらは主成分分析 (PCA) を用いた Dravnieks の匂い物質情報空間の解析を裏付けていると解釈できる。[068]

#### 1.10.1.4. RD Newcomb によるスーパー-Enose の提案

RD Newcomb らはキイロシヨウジョウバエ嗅覚の 60 種類の嗅覚受容体の報告された香気分子への応答をもとに、「電子鼻」が昆虫嗅覚バイオセンサーを採用した場合の識別能力についてシミュレーション検討をした[069]-[072]。識別回路として、人工ニューラルネットワーク (ANN) を介した、多層パーセプトロン (MLP) を採用した。生体の嗅覚系において、個々の匂い物質は、ORs のサブセットを活性化し、特定の ORs は、匂い物質のサブセットに対応するといったように、ORN 類は匂いコーディングの組合せのモデルを示すことが知られている。人工ニューラルネットワーク (ANN) は、コンビナトリアルなデータセット内に存在する匂いコーディングの本質的な特徴を学ぶことができる強力な適応フィルタであり、嗅覚認識の研究においても、データ内に存在する複雑な関係を同定する上で、ANN を利用することができる、としている。

まず、RD Newcomb らはアルコール、テルペン又はエステルといったそれらの化学的クラスのみに基づいて、電子鼻において信号処理のバックエンドとして使用されてきた、人工ニューラルネットワーク (ANN) を用いた場合の、昆虫嗅覚 e-nose の匂い物質の分類性能評価を検討した。受容体のすべてを使用した場合に対してにおい物質検出の 100% の性能を維持しながら大幅に 3 種類の受容体まで数を減らす場合にもバイオセンサの信号処理のバックエンドとしてどの程度うまく組み込むことができるのか、を検証した。

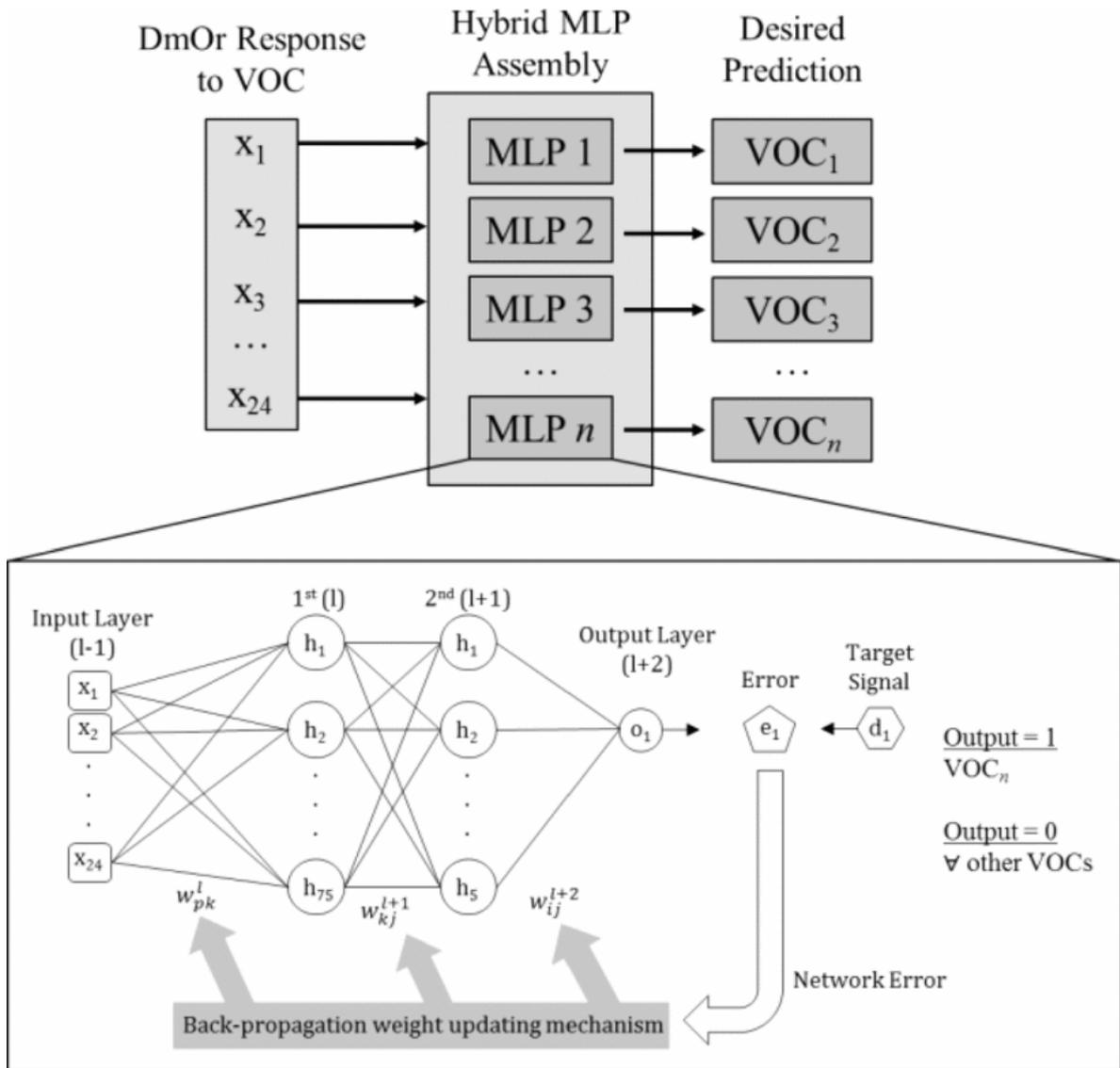


図 1-25. 複数の並列な多層パーセプトロン(MLP)の形をもつ人工ニューラルネットワーク [071]

次に、RD Newcombらは様々な濃度レベルで、そのような VOCs (揮発性有機化合物)や BVOCs (混合された揮発性有機化合物)を予測することが可能であるかを検証した。食品・飲料業界における電子鼻の潜在的な用途では、たとえば果実の成熟度の評価、ワインの品質や製品の細菌汚染の識別などのような、化学クラスを越えて予測でき、特定揮発性有機化合物として知られる化学物質や良い香りとして人々が感じる VOC および BVOCs の正確な同定ができるような能力が求められるためである。RD Newcombらは単一 VOCとBVOCsを予測するために、モデル生物ショウジョウバエの嗅覚受容体(DmOrs)の発火率の応答を複数の多層パーセプトロン(MLP)を使用したANNを使用することで分析し、ノイズ注入前後のデータからMLPの予測がどれくらい得られたのか、単一 VOC においてすべての VOC 検証ベクトルを同定した。

RD Newcombらはキロショウジョウバエの(ミバエ;DmOr)とハマダラカ(蚊;AgOr)、両方のモデル生物から ORs を導入した'スーパー-E-nose'を提案している。

現在の電子鼻やセンシングのフロントエンドとしてその採用昆虫嗅覚受容体(ORs)を利用した e-nose は、ORs が単一の種から来ているという事実によって潜在的に制限されている。キロショウジョウバエやハマダラカの 60 種類および 79 種類の嗅覚受容体を持つが、既存の「電子鼻」において昆虫嗅覚バイオセンサーを実装しようとする、受容体の機能発現やインターフェース実装の困難さから、検出のフロントエンドとして使用できそうな受容体の数は非常に低い。図 1-25 に示すような隠れ二層の多層パーセプトロン(MLP)複数個を並列にもつ人工ニューラルネットワーク(ANN)を用いることで、未知の匂い物質の化学クラスに対する識別にあって、最高の予測性能を提供できる最適な ORs を決定する手法について検討した。DmOr のみ、AgOr のみ、または DmOr-AgOr 種のクロス組み合わせからなる 3 種類の OR アレイについて、どれくらいすべての未知のにおい物質の検証セットを正しく分類したかを調査し、MLP を併用することで、未知の臭気物質の分類に使用可能な OR 種クロス‘スーパー Enose’を提案している。

この研究は昆虫の嗅覚受容体の応答データを使用しているが、識別能力の予測実験にとどまっている。

#### 1.10.1.5. ドイツのリューベック大学ママルーク

ドイツのリューベック大学で、嗅覚を定量化を試みた Madany ママルークは、自己組織化マップを用いて、嗅覚知覚空間におけるユークリッド近似の 2 次元マップを生成させた。[073]

ママルークらは、多次元尺度構成法(MDS)を使用した高次元のユークリッド空間へ嗅覚データベース(今回は知覚データベースである sigma Aldrich の flavor list [074]を使用)を投影し、匂い知覚の空間が計量的なのか非計量的なのか、を検討した。sigma Aldrich の嗅覚データベース flavor list をママルークは使用した。用いた知覚データベースは 851 種類の香気物質に対する官能評価による 278 種類の印象記述子からなる。これらのマップは、人間の匂い知覚の観点から、嗅覚の機能の潜在的な複雑さについて、新しい知識が引き出されることが期待される。また、嗅覚への低次元化のアプローチにより、匂い受容の情報空間に基づく匂いの構造的な基準を作ることができるとした。

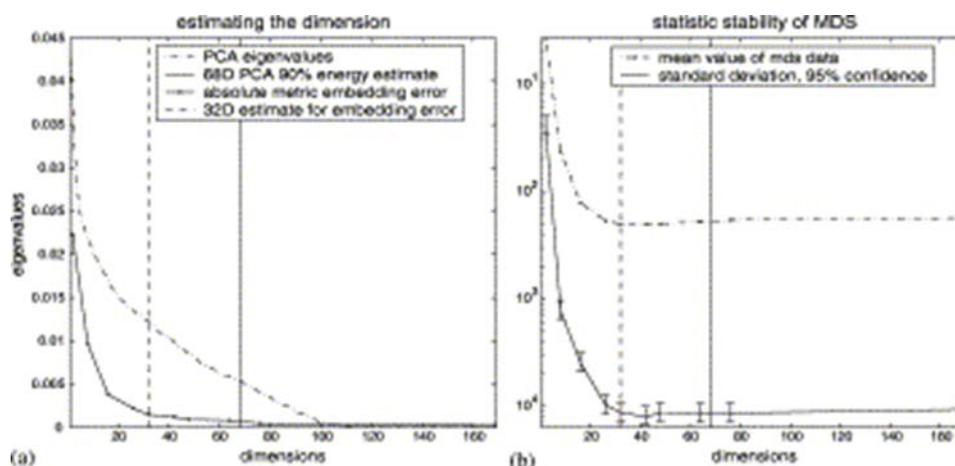


図 1-26. 次元の圧縮における誤差、MDS の安定性 ; (A) 次元数と誤差の関係、(B) 次元数と安定性関係 [073]

図 1-26(a)に、ママルークらの固有値分解による次元圧縮のプロットを示す。破線は 170 次元における固有値分解による次元圧縮を示し、図中の縦の線は、90%のエネルギーカットオフ(68 次元)を示す。誤差は、エネルギーグラフに合わせて拡大縮小した。32 次元以下になると次元における誤差(実線)が大幅に増加していることが分かった(垂直破線を参照)。図 1-26(b)に、MDS 標準偏差と MDS のシミュレーション実行の平均値を示す。スケールリングされた距離より値は、指定された非類似性に対応して減少した。

ヒトの嗅覚知覚は400種類前後の受容体に基づいているが、嗅覚知覚空間はそこまで高次元であることを意味するものではなかった、と今回の検討から結論された。彼らの本研究とチー・ローター[075]のデータを組み合わせ、結果、嗅覚知覚空間が少なくとも 32 次元の下限を持ち、68 次元の上限を持つという解釈が示された。またこれらの尺度の考察から、匂い物質の引き起こす神経活動についての(意味のある)予測と匂い物質単体の構造から匂い受容を導くことができると考察した。

## 1.10.2. マウスの嗅球マップデータに基づく検討

### 1.10.2.1. 嗅球活動の画像に基づく初段の嗅覚系の能力の分析

Jordi Fonollosa らは異なる匂いを検出し、識別するための情報工学的手法を適用して嗅球活動の画像から初段の嗅覚系の能力の分析を試みた。生体系の嗅覚受容体同士は低相関していることが示され、入力空間を幅広くコーディング可能であることを示した。彼女らは339種類の香気に対するマウスの嗅球の活動画像(1778ピクセル)を用いた。

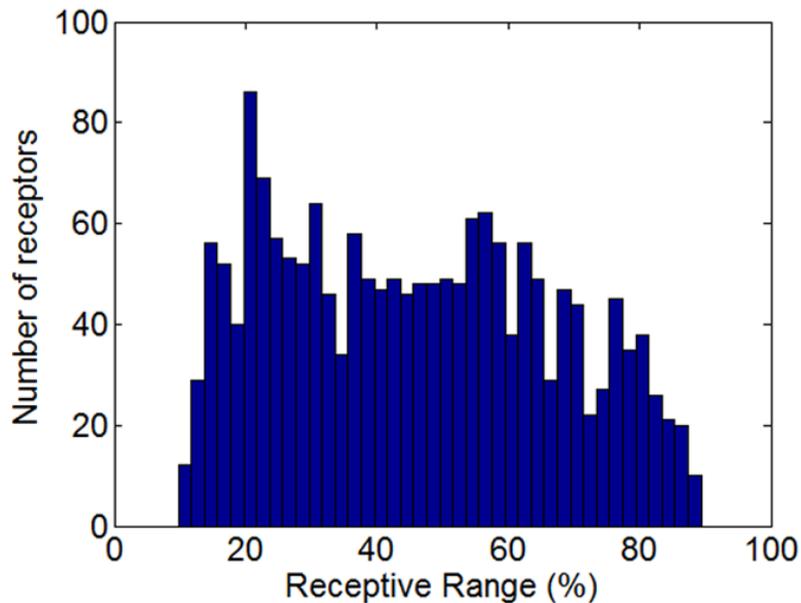
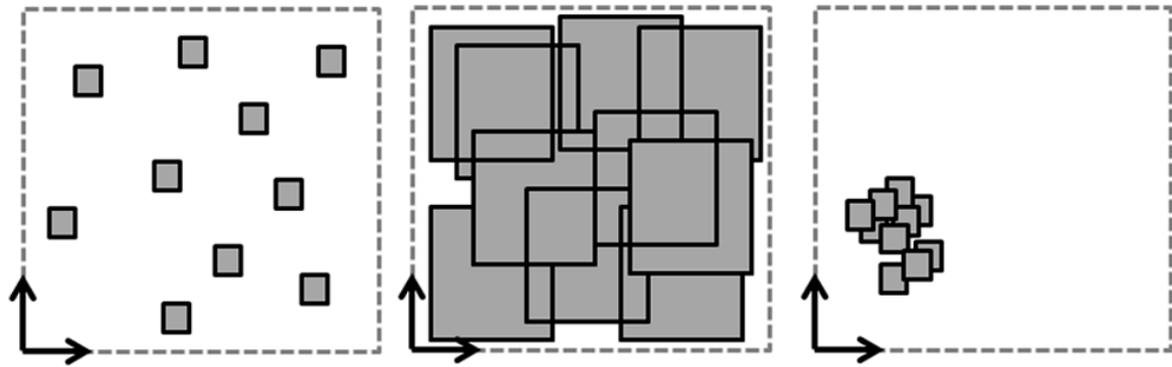


図 1-27. 複数の嗅覚受容体の相関関係とシステムの受容範囲選択性の相互関係 ; (A)抽象的な受容範囲、選択性と嗅覚受容体の相関関係、(B)1778 ピクセルの受容範囲の分布 [076]

図 1-27 に抽象的な受容範囲、選択性と嗅覚受容体の相関関係を示す。匂い空間と受容体のカバー領域を破線の正方形とそれぞれ黒い四角で表すとす。二つの異なる 10 種類の受容体のアレイ、それぞれ狭く～広く調整された受容体(左～中央)からなる、が作成されたとする。狭く調整された受容体は、あまり相関させることができないが、広く調整された受容体は匂いの空間の大きな領域をカバーし、より多くの匂い物質に対応できる。幅広いカバー領域を持つ受容体は、他の受容体との間のより多くのオーバーラップを持ち、相関させることができるが、小さなカバー領域の受容体も、右図のように相関させることはできる。図 1-27 下は嗅球の活動画像の各ピクセルが、どれくらいの匂い刺激に応答しているのか(受容範囲;receptive range)を調査した結果をヒストグラムに表わしたものである。狭く調整された受容体のみでなく、広く調整された受容体が組み合わせり、匂いの空間のより多くの匂い物質に対応できるであろうことが推察される。

Jordi Fonollosa らの研究から、生物学的または生体模倣であるかどうかによらず、汎用の化学センシングシステム的设计にはセンサの選択よりもセンサ間の低い相関が重要であることが結論された。[076]

### 1.10.3. 既存研究;各種センサ(MS, QCM etc) を用いた近似臭作成方法の検討

#### 1.10.3.1. これまでの近似臭作成方法

これまでに中本研究室は、匂いを自由に遠隔地で再現する装置・システムを目指して、質量分析器によるマススペクトル(Mass Spectrum:略して MS)データベースを用いた要素臭探索方法 1- 3)などを提案してきた。研究では質量分析データの測定にはガスクロマトグラフ質量分析器(GCMS-QP5050A, 島津製作所)を使用した。質量分析器で測定する際の試料の吸引と注入はオートインジェクタ(AOC-5000, 島津製作所)を用いることでたくさんの試料を連続して自動で測定した。

質量分析器で取得された各質量電荷比( $m/z$ ) における MS 強度は、香気に固有の 200 次元程度のベクトルと見なせることから、匂いの数値化が可能である。MS データは非負の値を持ち、その出力は線形重ねあわせが成り立つ。さらに、大量のデータを再現性良く得る上では質量分析器が優れているため、NMF 法を用いた要素臭の探索およびそれを用いた近似臭の作製の検討材料として好適である。

精油に関する匂い近似が出来ることを示されたが、質量分析器の広い出力強度範囲をカバーすることが匂いの近似精度向上に必要であることも明らかにされた。

#### 1.10.3.2. 近似臭と対象臭の高 $m/z$ 領域におけるマススペクトルの類似性

匂い物質の質量分析器によるデータベースに基づく NMF 法[080]を用いた多様な香りの近似手法の検討において、 $m/z$  の領域をいくつか区切って各領域におけるマススペクトルの類似性と三点識別法による官能検査結果による識別率との相関性が報告されている[078]。この中で高  $m/z$  領域(およそ  $m/z=150$  以上)における近似臭と対象臭の類似性との間には相関が見られたことが報告された。

中本研究室では過去に、研究室 11 人に対して 3 点識別法を行い、MS の特徴が人間の嗅覚に対してどのように現れるかを調査した[077]。135 種類の精油 MS データを元にした NMF 法による要素臭を作成し、得られた基底ベクトルを使用した 13 種類のブレンド精油の近似臭作成の結果を示す(基底数は 12 と 30 の 2 通り)。

図 1-28 に示すように低  $m/z$  領域における近似臭と対象臭の類似性との間には相関がほとんど見られないのに対し、高  $m/z$  領域における近似臭と対象臭の類似性との間には相関が見られ、高  $m/z$  領域の小さなピークの寄与を全体に反映させる必要性が示された。つまり、小さなピークであっても MS データ間の距離に十分な寄与をする尺度が必要である。[078], [079]。

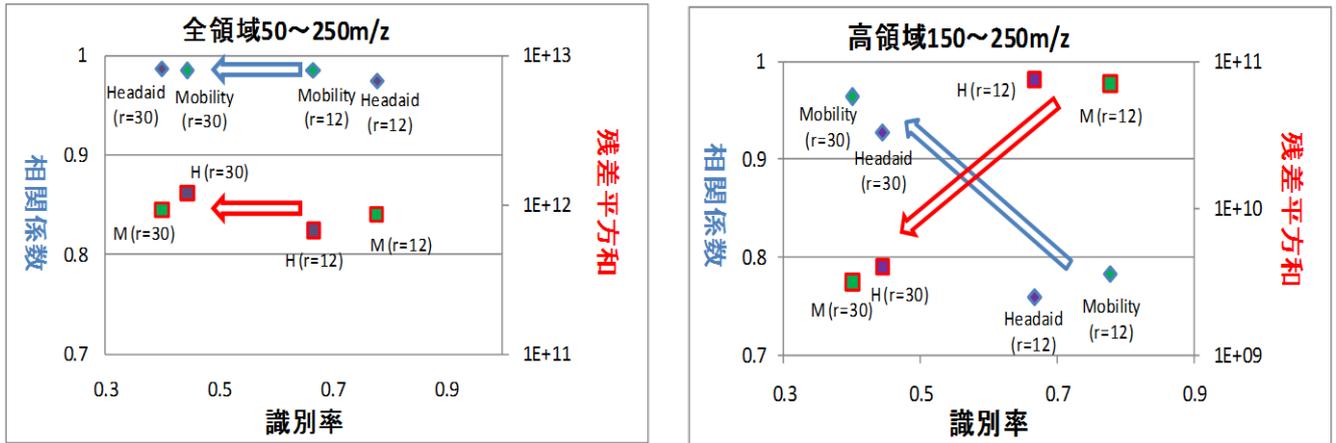


図 1-28. m/z の全領域(左)と高領域(右)領域におけるマススペクトルの類似性と、三点識別法による官能検査結果による識別率との相関性 [077]

だが、もともとのピーク強度の大きさが小さいために、ユークリッド距離のような通常用いられる距離尺度に基づく残差計算においては高 m/z ピークの残差の寄与は小さくなってしまい、小さいピークの多い高 m/z 領域のフィッティングが不十分になっていた。

## 1.11. 本研究の目的と論文構成

匂いは、人体において揮発性化合物の混合物が鼻の粘膜上の嗅覚神経細胞にて受容され、それによって生じる感覚刺激の対象である。

匂いを構成する香気物質は40万種類もあるとされ、その中には似たような印象のものも多い反面、微量で香気の影響に大きな影響を与える低閾値の香気物質も存在することが知られている。一部の化合物については、人間の感覚器によってガスクロマトグラフよりも鋭敏に検出される反面、分子構造が異なる化合物でも似た匂いとして感じることがあるこのように、生体の嗅覚システムは特定の物質群の差に対しては敏感だが別の物質群の差には鈍感だったりする。したがって揮発性の主に有機化合物の混合物である匂いの差異の評価において、‘構造’を人が感じているように捉えることと、単純な組成を測定する事とは、異なると考えられる。匂いを構成する香気物質の混合比率自体は、匂いの印象を表現するには冗長なデータであると考えられ、“感覚量”の空間による解釈が必要である。

本論文では、センサおよび分析器による手法、生理学的・分子生物学的手法、あるいは官能評価によって匂いが受ける多次元の符号化において、得られる情報空間を匂いセンシング空間と呼ぶことにする。

匂いはさまざまな生体内器官において異なる多次元データの形で情報として取り扱われている。もちろん匂いは生物によってのみ情報化されるわけではなく、センサアレイや分析機器によっても情報化することができる。用語と匂いの関係をトレーニングすることが必要だが、官能評価もヒトの知覚システム上において取り扱われる多次元情報といえる。

これまでに報告された多次元データ解析による匂い情報空間の解析は、心理学的なテストの情報空間を解析したものがほとんどであった。匂いセンシング空間の研究と、本論文で筆者が注目する研究対象の位置付けを図1-29にまとめる。

筆者らの研究室では、自由な遠隔地での匂い再現を目指して、匂いに関するセンシング空間・データ空間の解析を行ってきた。QCM センサを用いた匂い判別装置や、数百種類の匂い物質のマススペクトルから成るデータベースを構築し、マススペクトル空間上での基底ベクトルを非負値行列因子分解(NMF: Nonnegative Matrix Factorization)法を用いて抽出することによる、汎用的な役割を果たす要素臭の作成手法の提案を行った。基底ベクトルは実際に得られる香りのマススペクトルで近似したうえで、これらの調合比率を変えることで、多様な香りを近似臭として再現できることを報告した。しかし生体の嗅覚に近いセンシング情報を得るためには、より人間の感覚量空間に適したセンシング空間が必要になる点が指摘されてきた。

古くから匂いの要素が何なのか、というテーマは人々の関心を惹いてきた。嗅覚に高い関心が払われてきたことは、Dravnieksによる嗅覚に関する知覚データの収集などからも見て取れる。もしそれが決定できれば、複数の香りを任意の比率で混ぜ合わせて調合し、すぐに香りを発生させることができる。様々な香りを提示させられる嗅覚

ディスプレイを使用し、自由な遠隔地での匂い再現を目指した技術は実現に近づく。Jason B. Castro らの研究の研究を見てもわかるように、知覚面における、もしくは脳内における匂いセンシング空間を解析して、受容の次元を調査することが行われてきた。この分野に情報工学の果たす役割は大きい。

RD Newcomb らはキイロショウジョウバエの(ミバエ; DmOr)とハマダラカ(蚊; AgOr)、両方のモデル生物から ORs を導入した'スーパーE-nose'について検討した。彼らは隠れ二層の多層パーセプトロン(MLP)の形をもつ人工ニューラルネットワーク(ANN)を用いることで、未知の匂い物質の化学クラスに対する識別にとって、最高の予測性能を提供できる最適な ORs を決定する手法について検討し、OR アレイによる未知の臭気物質の分類に使用可能な OR 種クロス'スーパーEnose'の設計を提案した。

筆者は RD Newcomb らの研究にも注目している。この研究は昆虫の嗅覚受容体の応答データを使用しているが、識別能力の予測実験にとどまっているものの、嗅覚受容体による匂いセンシング空間に着目し、情報工学的アプローチで研究した事例であるためである。

本論文において、筆者は複数の香気物質に対する嗅覚神経細胞の応答情報を NMF 法等で解析することが、生体による最初段の“匂いコーディング(符号化)”を明らかにでき、匂いの情報化手法のための基礎的検討につながると考えた。本検討では、ヒトの嗅覚受容体の応答情報は利用できなかった。ヒトの嗅覚受容体の応答情報の取得は一定の成果を上げているが、ORN 全体の機能に関して検討できるほどの詳細な機能はまだ報告されていない。

以上を踏まえて本研究では、(昆虫の)嗅覚受容体による匂いセンシング空間をもとにした要素臭探索法の開発を目的として、二点の観点から検討を行った。

- 嗅覚受容体による匂いセンシング空間の解析方法
- 嗅覚受容体による匂いセンシング空間の補間方法

ORN による匂いセンシング空間としては、比較的応答情報が得られている昆虫のデータを用いることにした。各章の関係は図 1-30 のようになる。

	匂いセンシング	ヒト・哺乳類	知覚・心理学	高次脳機能	嗅球	受容体・ORN	嗅球	受容体・ORN	解析手法	結論
	非生物						昆虫		アルゴリズム名	研究の目的・主に行なった事・成果
研究者名・論文名									NMF	要素臭と近似集作成手法
中本研究室	MS								NMF	匂い識別実装
中本研究室	QCM								LVQ	匂いにおけるホロイトノイズの定義(+官能評価実験)
Noam Sobel			Dravnieks "ATLAS of.."						PCA	描述子空間の146記述子のクラスタ化、匂い物質空間の144個の匂いクラスタ化→匂い空間は連続的な知覚ではなく、クラスターに分けることができる
Jason B. Castro			Dravnieks "ATLAS of.."						NMF	匂いの認識空間の主成分分析と匂いへの嗜好性の関係(分子情報dragon併用)
Rehan M. Khan			Dravnieks "ATLAS of.."						PCA	知覚レベルの嗅覚の機能の次元数探索
Amir Madany Mamlouk			sigma aldrich 'favor list'						多次元尺度法	嗅球活動の画像に基づく初段の嗅覚系の能力の分析、化学センシングシステム的设计にはセンサの選択よりもセンサ間の低い相関が重要
Jordi Fonollosa									--	要素臭探索
本研究									NMF	
RD Newcomb									MLP(ANN)	昆虫嗅覚バイオセンサーを採用した'E-nose'の識別能力シミュレーション
(備考) 工学・生物学的知見										
代表的な手法	GC, LC, MS, ガスセンサ (QCM, 無機半導体, 有機半導体センサ)	官能検査 (ヒトのみ)	fMRI, NIRSなど	電気的手法など多様	細胞染色、ヘテロ発現細胞利用など多様	(未調査)	電気的手法、ヘテロ発現細胞利用など多様			
大規模データの利用	◎	△	×	△	×~△	○~△	○~△			
後天的な機能チューニング (生物)	無 (再現性良)	有	有	有 (発達期)	ほぼ無	(未調査)	ほぼ無			
筆者注		評価用語は匂いトレーニングによって獲得される	匂い学習によって脳の働き方が変化する							
*解説章		1.9.6.1.	1.9.6.2.	1.9.6.3.	1.7. 1.9.7.5. 1.9.7.8.	1.9.6.4.	1.7. 1.9.7.3. 1.9.7.7.			

図 1-29. 匂いセンシング空間の研究の種類

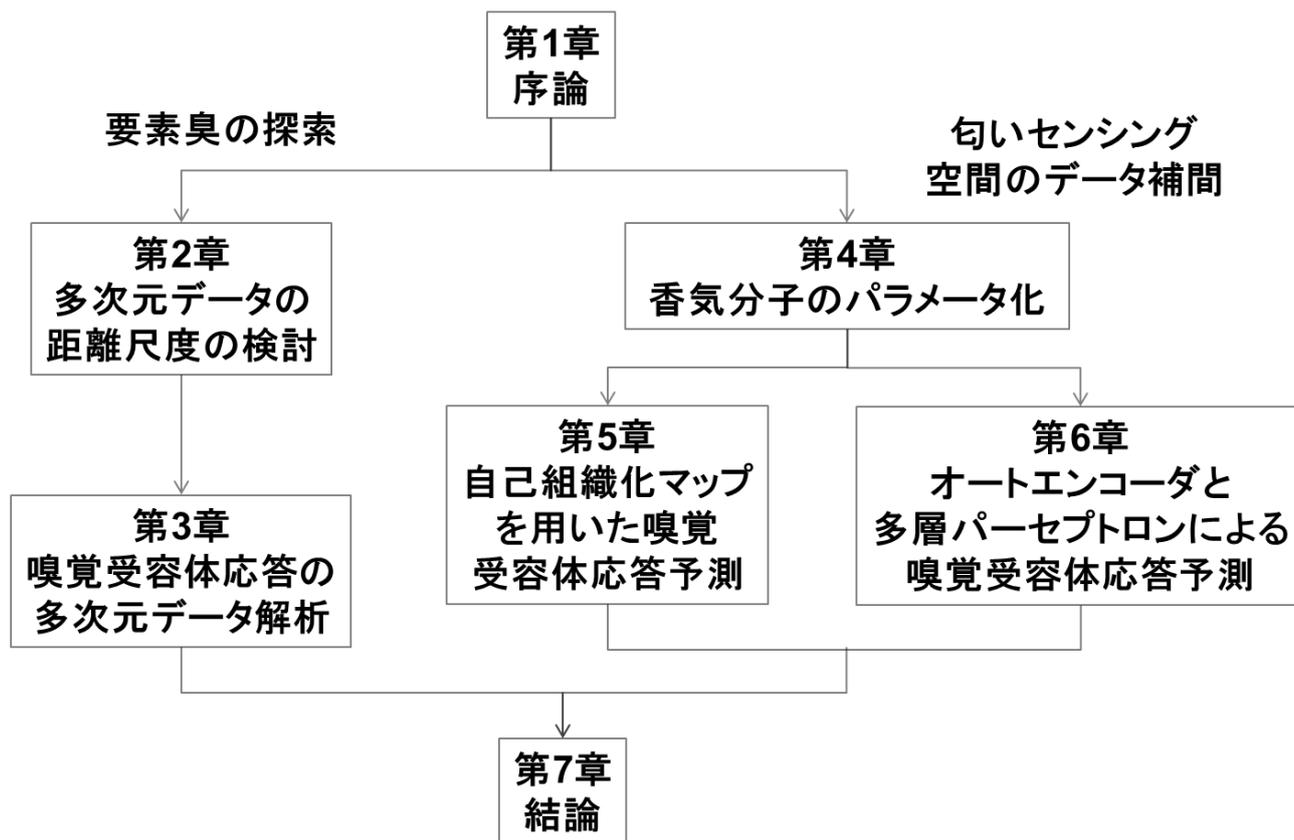


図 1-30. 各章の関係

上記一点目の観点は、具体的には嗅覚受容体による匂いセンシング空間に基づく、最適な距離尺度の選択と NMF 法による要素臭の探索を行った。筆者は、まず第二章にて香気データ間の距離尺度に注目した。広いダイナミックレンジを得るのに最適な距離尺度を検討について説明する。次に第三章で、人間の感覚量空間に適したセンシング空間が必要との指摘に対し、昆虫の嗅覚受容体の応答データの解析に関して説明する。最適な距離尺度の選択し、NMF 法などを用いて検討することで、嗅覚受容体による匂いセンシング空間から要素臭を探索する。

本検討では、人の嗅覚受容体の応答情報は利用できなかった。ヒトの嗅覚受容体の応答情報の取得は一定の成果を上げているが、ORN 全体の機能に関して検討できるほどの詳細な機能はまだ報告されていないので、比較的応答情報が得られている昆虫のデータを用いることにした。

上記二点目の観点は、具体的には香気物質の分子構造をもとにした嗅覚受容体の応答強度予測を検討した。昆虫の ORN 応答データベースでも限られた化合物に対する応答データしか得られていないことから、受容体の応答予測が必要となった。第四章から第六章で、筆者はより多くの香気物質に対して分子構造を考慮した受容体の応答予測の構築するため、香気分子の分子構造パラメータに基づく自己組織化写像(SOM)および多層パーセプトロンを用いた、嗅覚受容体の応答強度を予測する手法を検討する。

昆虫の嗅覚受容体の応答データを NMF 法を用いて解析し、実用的な要素臭が探索できる可能性を調査した検

討は、他に例はない。筆者の研究は、ヒトによる最初段の“匂いコーディング(符号化)”の解析、匂いの情報化手法につながると考えられる。また生物学的な受容体の機能解明データの蓄積は近年急速に進んでいる反面、匂いを構成する香気物質に対する、すべての嗅覚受容体の応答を取得することは事実上不可能であり、使用可能な香気応答データは限定されている。本論文にて報告する予測手法は基礎検討的であり、改善すべき点も多い。だが、分子構造を元にした情報空間を利用することで、多種多様に存在する香気物質に対する匂いセンシング空間との非線形なマッピングを得、匂いセンシング空間を補間する手法は、匂いセンシング空間の解析に貢献できるものと考えている。

## 第二章 非負値行列因子分解法を用いた近似臭作成方法における距離尺度の検討

香りの情報化には、対象とする香りを表現する少数の要素臭の確立が不可欠である。異なる香りを混ぜて嗅ぐと異なるイメージの香りが作り出せる。要素臭を確立することにより、図 2-01 に示すような手法で香りの再現が容易にできるようになると考えられる。

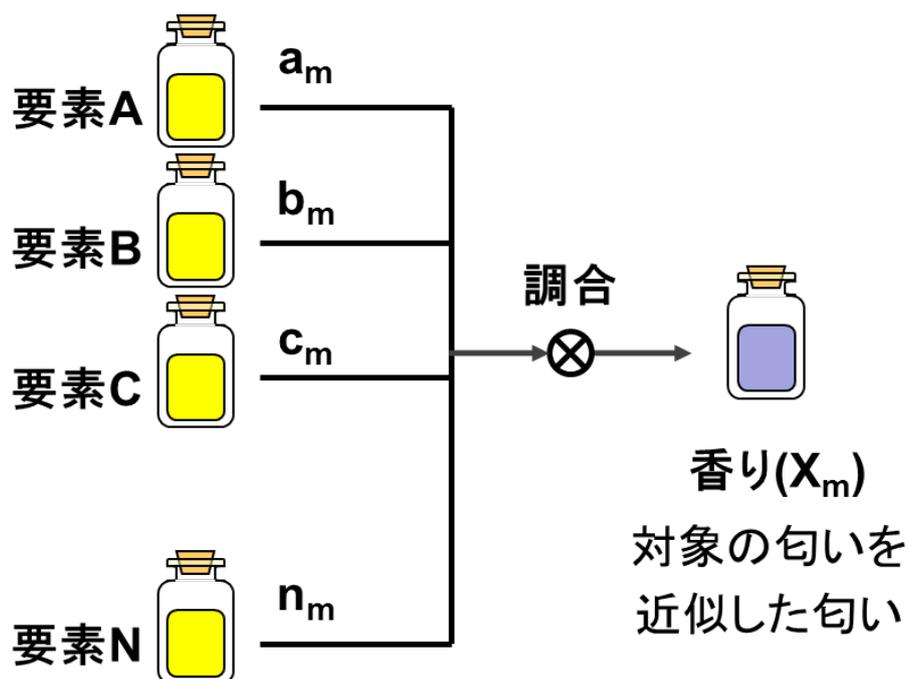


図 2-01. 香りの再現手法

香りの基本要素の探索に用いることができる官能検査に基づく膨大なデータは、得るのは困難である。そのため汎用的な要素臭はまだ明らかになっておらず、匂いの情報を元に、その香りを要素臭とレシピを用いて調合する方法を開発は報告されていなかった。

これまでに中本研究室は、匂いを自由に遠隔地で再現する装置・システムを目指して、質量分析器によるマススペクトル(Mass Spectrum:略して MS)データベースを用いた要素臭探索方法[077]-[079]などを提案してきた。そこで精油に関する匂い近似が出来ることを示したが、質量分析器の広い出力強度範囲をカバーすることが匂いの近似精度向上に必要であることも明らかになった。

そこで本研究では香気データ間の距離尺度に注目し、広いダイナミックレンジを得るのに最適な距離尺度を検討した。なおここで、広いダイナミックレンジと称しているのは、強い信号強度から微小な信号強度まで広範囲の強度を持つデータの再現性の良さを表している。広いダイナミックレンジを検討することになった経緯については 2.2.3 節で説明をする。

## 2.1 データベースとその基底について

$$\begin{array}{cccc}
 \text{Database} & & \text{Basis vectors} & & \text{coefficients} & & \text{Approximated matrix} \\
 \begin{bmatrix} v_{1,1} & \cdots & v_{1,m} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,m} \end{bmatrix} & \cong & \begin{bmatrix} w_{1,1} & \cdots & w_{1,r} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,r} \end{bmatrix} & \times & \begin{bmatrix} h_{1,1} & \cdots & h_{1,m} \\ \vdots & \ddots & \vdots \\ h_{r,1} & \cdots & h_{r,m} \end{bmatrix} & = & \begin{bmatrix} wh_{1,1} & \cdots & wh_{1,m} \\ \vdots & \ddots & \vdots \\ wh_{n,1} & \cdots & wh_{n,m} \end{bmatrix} \\
 V & & W & & H & & WH \\
 (n \times m) & & (n \times r) & & (r \times m) & & (n \times m)
 \end{array}$$

図 2-02. 香りのデータベース V と要素臭の関係

香りのデータベース V と要素臭の関係は、図 2-02 のように表現できる。できるだけ直交するように抽出された基底行列 W と、係数行列 H との積として、データ空間全体 V が近似できるとする。ここで図 2-02 中の n は香気データの次元数、m はデータベースに含まれる香気データのサンプル数、r は基底数である。係数行列 H の各列ベクトルはデータ V を近似するための各基底ベクトルの構成比と考えることができる。要素臭を各係数どおりに調合し、各香気と同じ特徴の匂いを近似できる。香りデータの分解、基底行列 W の確立のためには、様々な香りのデータを取得しデータベース化し、次元の圧縮をすることで行える。

香りのデータベース V を構築する上で、生体の嗅覚に近いセンシング情報を得ることが望ましく、官能検査のデータを用いることや、嗅覚受容体の応答に基づいた空間を探索することが有効だと考えられる。前者において膨大かつ信頼性を持つ官能検査のデータを得るのは困難である。なお生体の嗅覚情報については、現在幅広い知見が蓄積されつつあることをうけ、後者において、筆者らは嗅覚受容体の応答に基づいた空間を探索することで近似臭作成方法を提案している[081]。次元の圧縮方法として代表的な手法を 2 例、次項において紹介する。

### 2.1.1 主成分分析 (PCA)

主成分分析 (Principal Component Analysis, PCA) は、直交回転を用いて変数間に相関がある元の観測値を、相関の無い主成分とよばれる値に変換する手法である。[082]

要素が非負とは限らない系に適用が可能である。主成分分析により、多次元のデータを情報量をなるべく落とさずに低次元に要約することができる。元の観測値 (対象) に対応した変換後の値である主成分得点と、各々の主成分得点に対する変数の重みに相当する主成分負荷量として主成分分析の結果は得られ、寄与率が高い少数個の主成分を用いて元の観測値を表現することは、次元を圧縮することに相当する。

ある主成分にデータ全体の情報がどれぐらい含まれているかは、その主成分に対応する固有値 (分散) が固有値和 (分散の和) の中でどれぐらいの割合を占めているかで説明できる。各固有値が固有値全体に占める割合を寄与率、その寄与率を累積したものを累積寄与率と呼ぶ。主成分分析は広く知られ、多く用いられる多次元データ解析方法の一つであり、1.10 節で紹介した既存研究でも使用されている。

### 2.1.2 非負値行列因子分解法 (NMF)

NMF 法はある行列  $\mathbf{V}$  が存在した場合に、図 2-03 に示すような近似行列を導く計算手法である[080]。すべての要素が非負である拘束付きアルゴリズムであり、データ行列も得られる基底行列も非負の値のみから成る。 $n$  次元ベクトルの形を持つデータ  $m$  個をデータ行列  $\mathbf{V}(n \times m)$  の行列とする。データ行列  $\mathbf{V}$  に NMF 法を実行するには、行列  $\mathbf{W}, \mathbf{H}$  に初期値を与え、更新しながら繰り返し計算を行う。 $\mathbf{V}$  と  $\mathbf{WH}$  の距離を最小化していくことで、 $\mathbf{W}$  と  $\mathbf{H}$  を得る。この計算により、基底行列  $\mathbf{W}$  と係数行列  $\mathbf{H}$  の積に分解され、近似行列  $\mathbf{WH}$  の各列ベクトルは元のデータに対する近似データである。基底ベクトル数  $r$  の値を  $m$  より小さい値に設定するので  $\mathbf{W}$  は  $\mathbf{V}$  の次元の圧縮したものに相当する。

今回のように要素臭を検討したい場合、非負の係数が導出されることが望ましい。PCA は直接使いにくいものの、NMF であれば非負の係数が導出されるので適している。

(a) 要素臭作成方法(NMF method etc)

Database	Basis vectors	coefficients	Approximated matrix
$\begin{bmatrix} v_{1,1} & \cdots & v_{1,m} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,m} \end{bmatrix}$	$\cong \begin{bmatrix} w_{1,1} & \cdots & w_{1,r} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,r} \end{bmatrix}$	$\times \begin{bmatrix} h_{1,1} & \cdots & h_{1,m} \\ \vdots & \ddots & \vdots \\ h_{r,1} & \cdots & h_{r,m} \end{bmatrix}$	$= \begin{bmatrix} wh_{1,1} & \cdots & wh_{1,m} \\ \vdots & \ddots & \vdots \\ wh_{n,1} & \cdots & wh_{n,m} \end{bmatrix}$
$\mathbf{V}$	$\mathbf{W}$	$\mathbf{H}$	$\mathbf{WH}$
$(n \times m)$	$(n \times r)$	$(r \times m)$	$(n \times m)$

(b) 近似臭作成方法 \*現実的にはWは近似基底行列を用いる

Target odor	Basis vectors	Recipe	
$\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$	$\cong \begin{bmatrix} w_{1,1} & \cdots & w_{1,r} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,r} \end{bmatrix}$	$\times \begin{bmatrix} s_1 \\ \vdots \\ s_r \end{bmatrix}$	$= \begin{bmatrix} ws_1 \\ \vdots \\ ws_r \end{bmatrix}$
$\mathbf{t}$	$\mathbf{W}$	$\mathbf{s}$	$\mathbf{ws}$
$(n)$	$(n \times r)$	$(r)$	$(n)$

図 2-03. NMF 法による質量分析器を用いた近似臭作成方法

得られた基底ベクトルに相当する匂いが自然界に存在するとは限らないため、二段階の近似計算を行う。最初の近似計算は基底ベクトルの既存の匂いによる近似、この近似により得られた近似基底行列を用いて、図 2-03b に示すように、ターゲット香気データを近似した。近似すべきターゲット香気データ  $\mathbf{t}$  を、基底行列  $\mathbf{W}$  とレシピ  $\mathbf{s}$  の積  $\mathbf{ws}$  で近似する時、 $\mathbf{t} \sim \mathbf{ws}$  間の距離が最小となるような  $\mathbf{s}$  を探索するために、非負拘束付最小二乗法などを用いる必要がある。得られた  $\mathbf{s}$  は要素臭を調合する際のレシピに相当する。

## 2.2 マススペクトルのデータにもとづく NMF 法を用いた要素臭の探索

### 2.2.1 マススペクトルのデータの特徴と本研究で使用したデータ

当研究室では、匂いのデータベースを構築し、非負値行列因子分解 (NMF) 法を用い、汎用的な役割を果たす要素臭を抽出、それを用いた多様な香りの近似について報告してきた[077]-[079]。図 2-04 に示すように、質量分

析器で取得された各質量電荷比(m/z) におけるMS 強度(マススペクトル)は、香気に固有の 200 次元程度のベクトルと見なせることから、匂いの数値化に使用できる。MS データは非負の値を持ち、その出力は線形重ねあわせが成り立つ。さらに、大量のデータを再現性良く得る上では質量分析器が優れているため、NMF 法を用いた要素臭の探索およびそれを用いた近似臭の作製の検討材料として好適である。

### 2.2.2 マススペクトルのデータ取得方法

これまで質量分析(MS)によるデータ空間上で NMF 法を適用することによって、汎用的な役割を果たす要素臭の作成し、近似臭の作成を検討した。本章で使用するマススペクトルデータは、本研究室で以前に取得されたデータである。データの取得方法は 2012 年卒業の二瓶氏[083]や 2010 年度卒業村上氏[084]、2013 年度卒業本田氏[085]の手法に準じ、表 2-01 に示す条件で行った。以下に要点をまとめる。

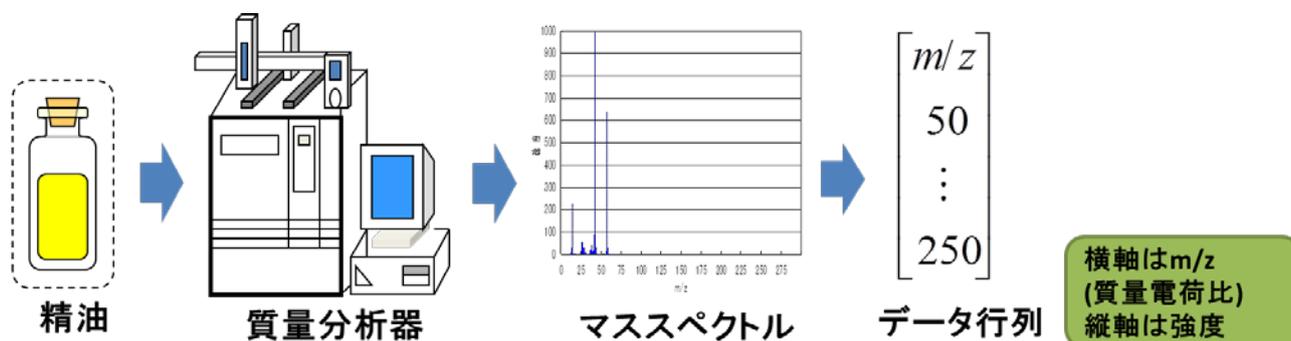


図 2-04. 質量分析により得られるデータ

表 2-01. 質量分析器を用いた実験条件

質量分析器の仕様		測定パラメータ	
質量範囲	10～900m/z	注入量	1μL
イオン化方式	EI法(70eV：固定)	スプリット比	156
質量分離部	四重極型	インジェクション温度	250℃
スプリット/スプリットレス注入ユニット		カラム温度	300℃
カラム内径	0.10mm	ディテクター温度	300℃
カラム長	10m	サンプリング時間	3min
カラムコーティング	なし	全流量	11 ml / min
キャリアガス	ヘリウム 純度99.995%以上		

質量分析データの測定にはガスクロマトグラフ質量分析器(GCMS-QP5050A, 島津製作所)を使用した。質量分析器で測定する際の試料の吸引と注入はオートインジェクタ(AOC-5000, 島津製作所)を用いることでたくさんの

試料を連続して自動で測定した。

また雑音除去法について検討したうえで、以下の条件でデータ前処理は行うこととした。質量分析器でエタノール(分子量:46.07)を連続50回測定し、雑音成分となる質量電荷比(m/z)50~250における各m/zの強度を調べ、閾値=平均+標準偏差×1.5よりも小さな値を各m/zについてはノイズであると判断し、0にする前処理を行った。

### 2.2.3 中本研究室での質量分析(MS)による要素臭の作成と近似臭の作成の検討

中本研究室では、表 2-02 に示す 135 種類の精油について検討し、NMF 法における距離尺度としての IS 距離の有用性と、多様な香りの近似について報告した。MS データ全体(香気物質 135 種類)のデータに、ユークリッド距離、KL ダイバージェンス、IS ダイバージェンスの 3 種類を距離尺度とした NMF 法を適用し、得られたデータを比較、検証した。

**表 2-02. 135 種類の精油データを取得し、NMF 法で要素臭抽出 [084], [085]**

No.	精油	No.	精油	No.	精油	No.	精油	No.	精油
No.1	Breathesay	No.28	CELERY SEED	No.55	Lavender	No.82	Sw_marjoram	No.109	タナセタム
No.2	Daybreak	No.29	Chamomile_G	No.56	Lemon	No.83	Sw_orange	No.110	タラゴン
No.3	De-stress	No.30	Chamomile_M	No.57	Lemongrass	No.84	Tea tree	No.111	ディール
No.4	Equilibrium	No.31	Chamomile_R	No.58	Lime	No.85	Thyme	No.112	ナルシス
No.5	FOCUS	No.32	Cinnamon_L	No.59	Mandarin	No.86	Valerian	No.113	ハニーサック
No.6	Goodnight	No.33	Citronella	No.60	MANUKA	No.87	Vetiver	No.114	バルサムモミ
No.7	Headaid	No.34	Clary sage	No.61	May chang	No.88	Yarrow	No.115	ヒノキ
No.8	MOBILITY	No.35	Clove bud	No.62	Melissa	No.89	Ylang Ylang	No.116	ヒマラヤスギ
No.9	Noel	No.36	Coriander	No.63	Myrtle	No.90	Ceder wood	No.117	ヒヤシンス
No.10	Refresh	No.37	Cypress	No.64	Neroli	No.91	Litsea	No.118	ブラックスブルース
No.11	Relaxation	No.38	ELEMI	No.65	Niaouli	No.92	Rosewood	No.119	ヘリクリサム
No.12	Romance	No.39	Eucalyptus_C	No.66	NUTMEG	No.93	Winter green	No.120	ベルカモットミント
No.13	Travel	No.40	Eucalyptus_G	No.67	Organic orange	No.94	アジヨワン	No.121	ボンテローザパイン
No.14	Angelica root	No.41	Eucalyptus_R	No.68	Palmarosa	No.95	アトラスシダー	No.122	マスティックウリー
No.15	Basil	No.42	Fennel_S	No.69	Patchouli	No.96	アプリコット	No.123	モミ
No.16	BAY	No.43	FIR (Siberian)	No.70	Peppermint	No.97	アルペンシスミント	No.124	ユズ
No.17	Benzoin	No.44	FIR BALSAM	No.71	Petitgrain	No.98	オークモス	No.125	ヨーロッパアカマツ
No.18	Bergamot	No.45	Frankincense	No.72	Pine	No.99	オレガノ	No.126	ラブラドルティ
No.19	Black pepper	No.46	GALBANUM	No.73	RABENSARA	No.100	カラミント	No.127	ランタナ
No.20	Cajeput	No.47	Geranium	No.74	Rose absolute	No.101	クエラ	No.128	ローレル
No.21	CAMPHOR	No.48	Ginger	No.75	Rose otto	No.102	シソ	No.129	ロックローズ
No.22	CARAWAY	No.49	Grapefruit	No.76	Rosemary	No.103	シナモスマフラグランス	No.130	ワイルドキャロット
No.23	Cardamon	No.50	Helichrysum	No.77	S_majororam	No.104	スターアニス	No.131	イリス
No.24	Carrot Seed	No.51	Ho Wood	No.78	Sage_S	No.105	スパイクナード	No.132	キンモクセイ
No.25	Cedarwood_A	No.52	Jasmine	No.79	Sandalwood	No.106	セロリ	No.133	トウルールバルサム
No.26	Cedarwood_H	No.53	Juniperberry	No.80	Spearmint	No.107	ダヴァナ	No.134	ミモザ
No.27	Cedarwood_V	No.54	Lavandine	No.81	Spikenard	No.108	タジェティーズ	No.135	ミユラ

第一章でこれらの基底ベクトル抽出と要素臭の検討については述べたが、ピークが小さい高  $m/z$  領域の再現性は低かった。当研究室では、これまでに各  $m/z$  のピークに重み付けをする方法や各強度の対数をとる方法が検討されたが、満足できるような結果は得られていなかった。匂いの特徴をより再現するために、NMF法においてダイナミックレンジが広い距離尺度を用いることが有用と考えられる[078], [079]。NMF法におけるデータ行列  $V$  ~ 近似行列  $WH$  間の距離尺度としては、様々な距離尺度に従った更新式が報告されている[086]。

## 2.3 NMF 法における距離尺度

### ユークリッド距離

$$D_{EU}(V|WH) = \sqrt{\sum_{ij} (V_{ij} - WH_{ij})^2} \quad (1)$$

### KLダイバージェンス

$$D_{KL}(V|WH) = \sum_{ij} \left( V_{ij} \log \frac{V_{ij}}{WH_{ij}} - V_{ij} + WH_{ij} \right) \quad (2)$$

### ISダイバージェンス

$$D_{IS}(V|WH) = \sum_{ij} \left( \frac{V_{ij}}{WH_{ij}} - \log \frac{V_{ij}}{WH_{ij}} - 1 \right) \quad (3)$$

図 2-06. 多次元の場合、各距離関数の定義

高  $m/z$  領域における近似臭と対象臭の類似性との間には相関が見られ、高  $m/z$  領域の小さなピークであっても MS データ間の距離に十分な寄与をする尺度が必要である。匂いの特徴をより再現するために、NMF法においてダイナミックレンジが広い距離尺度を用いることが有用と考えられる[078], [079]。筆者は、距離尺度としてダイナミックレンジが広い板倉斎藤(IS)距離を導入した NMF法を用いることで、高  $m/z$  領域に低強度な MS ピークとして出現している匂いの特徴を再現できるか検証した。

多次元の場合、各距離関数の定義について図 2-06 にまとめた。ユークリッド距離は式(1)のように、Kullback-Leibler (KL)ダイバージェンスは式(2)のように、板倉-斎藤 (IS)ダイバージェンスは式(3)のように、距離は書き表せる。

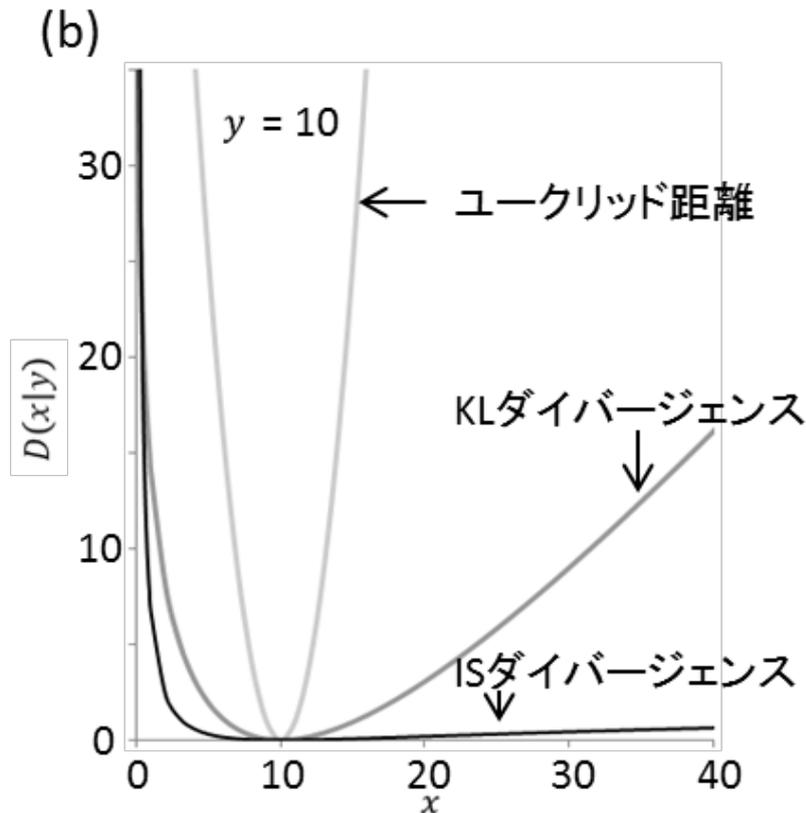


図 2-07. 一次元の場合の各距離関数の定義に基づく  $x \sim y$  間の距離 ( $y=10$ )

一次元の場合の各距離関数の定義に基づく  $x \sim y$  間の距離を  $y = 10$  に固定して、 $x$  の関数として図 2-07 に示す。図 2-07 から分かるように  $x > 10$  の領域において、ユークリッド距離、KL ダイバージェンスにおいては、 $x \sim y$  間の間隔が広がると距離の大きな増大がみられるのに対して IS ダイバージェンスにおいては距離増大は小さくなる。 $x < 10$  の領域では、 $x \sim y$  間の距離が相対的に大きく捉えられる。狭い間隔の  $x \sim y$  間が相対的に大きな距離として捉えられ、ダイナミックレンジが広いデータの場合に IS ダイバージェンスは有効である。

IS ダイバージェンスの大きな特徴は、行列  $\mathbf{V}$  の値のダイナミックレンジが広い状況でも、データ強度の大きい値と小さい値を同程度に扱うことといえる。匂いの特徴が現れる高領域  $m/z$  の再現性を高めるためには、NMF 法における距離尺度として IS ダイバージェンスが適していると考えられた。なお図 2-07 は説明のために一次元における距離尺度を図示したが、実際には質量分析器データの場合、200 次元程度である。

図 2-08 にさまざまな距離尺度を持つ NMF 法計算を示す。NMF 法におけるデータ行列  $\mathbf{V}$  ~ 近似行列  $\mathbf{WH}$  間の距離尺度としては、様々な距離尺度に従った更新式が報告されている[086]。

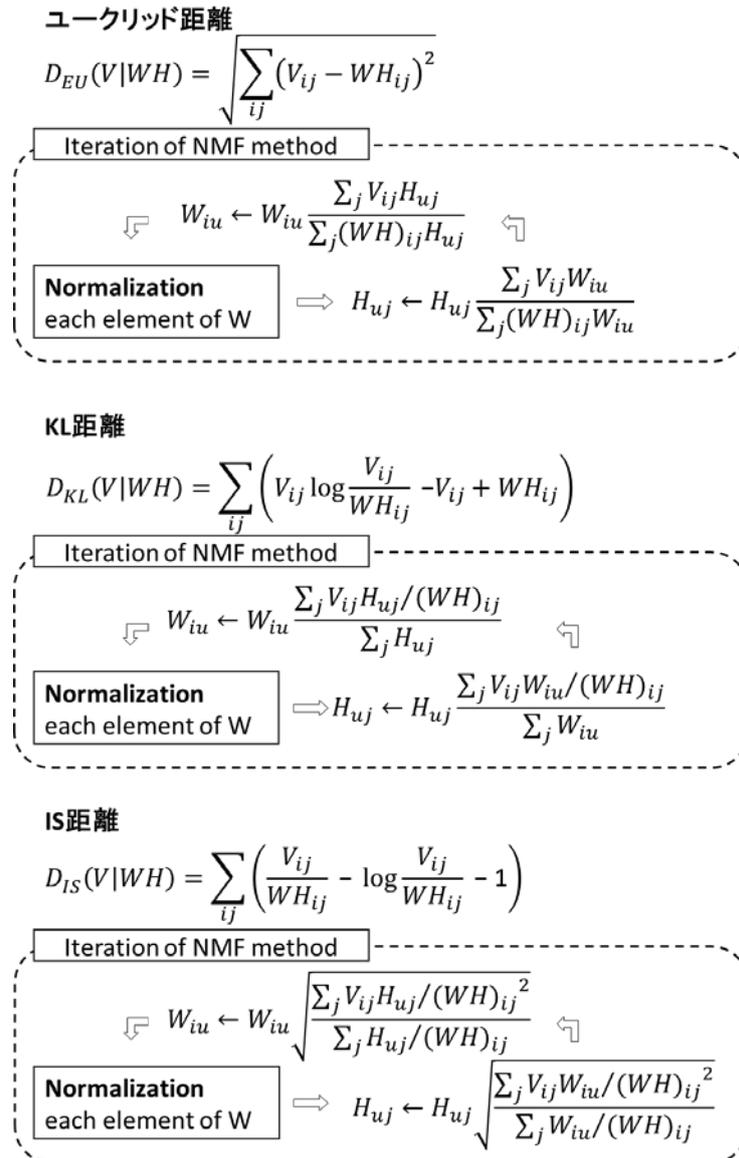


図 2-08. さまざまな距離尺度を持つ NMF 法計算

各距離尺度とも通常は乱数にて初期値を与え、ゼロによる除算を防ぐために微小な値を加えた。マススペクトルに対する NMF 法の適用において、各サンプルのマススペクトルが同じになるように  $\mathbf{V}$  の各列の総和が一致するように規格化した。同じ目的で、初期  $\mathbf{W}$  では初期値を 0 から 1 までの乱数にて与え、各マススペクトル強度が同じになるように初期  $\mathbf{W}$  の各列の総和が一致するように規格化した。繰返し計算によって  $\mathbf{W}, \mathbf{H}$  を得る。主成分分析とは異なり、第一主成分、第二主成分、第三主成分などが順位も含め一意的に計算により決定するものではなく、必ずしも直交性が保証されるわけではないことが NMF 法の利用において留意すべき点となっている。

## 2.4 近似レシピ計算における IS 距離の導入と MATLAB 上実装

これまで、大規模な質量分析に基づく香気のパターンデータベースより NMF 法によって得られた基底の再現において、要素臭のレシピを基底抽出に用いた香気を重ね合わせにより再現する検討を重ねてきた。この要素臭レシピを決定する際、最小二乗法に基づいて、誤差が最小になるような各香気と、その香気の構成比(処方中配合量)を

MATLAB 上で計算してきた。

これまで NMF 法における距離尺度としてユークリッド距離および KL 距離を用いて NMF 法の適用を検討してきたが、NMF 法における距離尺度として IS 距離の使用を新たに検討し、報告した。KL 距離および IS 距離はユークリッド距離とは異なる距離尺度である。このような距離尺度を用いて NMF 法より得られた基底ベクトルの近似法として、ユークリッド距離に基づく最小二乗法を用いることに対して適用することの正当性に疑問があった。

このような基底を再現する構成比の決定に IS 距離もしくは KL 距離を導入できる方法として、MATLAB 上における `fmincon` 関数がある。任意の距離関数を定義し、その関数が返す値を指定値に近づける(今回の場合では最小となる、本来は最大値、最小値、もしくは任意の値)に近づけるパラメータを返すコマンドである。これを使用することでユークリッド距離における最小二乗法ではなく、任意の距離計算式に基づく極小点を求めることができる。

今回使用する構文 ; `x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon, options)`

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases}$$

この MATLAB 関数は指定された問題の最小値を見つける。`fun` はベクトル `x` を使って `x` で計算した目的関数のスカラー値 `f(x)`を返し、最小化される関数である。多様な制約も可能(以下に簡易に説明)だが、本研究では `lb = 0` (ベクトル `x` のすべての要素) の制約のもと実行した。

設計変数 `x` に上限と下限を定義されており、解が常に `lb ≤ x ≤ ub` の範囲に存在するように、`x0` を探索開始点として計算される。等式が存在しない場合には `Aeq = []` と `beq = []` を設定する必要がある。`A` と `Aeq` は線形制約行列、`b` と `beq` はそれらに対応したベクトルであり、`Aeq*x = beq` と `A*x ≤ b` を制約する。`c(x)` と `ceq(x)` はベクトルを返す関数で、`nonlcon` で定義された非線形不等式 `c(x) ≤ 0` または等式 `ceq(x) = 0` の下で制約して最適化するが、非線形不等式制約または非線形等式制約がない場合には `nonlcon = []` を設定する。`options = optimoptions('fmincon','GradObj','on')` のように指定することで `fun` の勾配を計算することもできる。収束点探索における探索方向ベクトルを定義して使用することで、効率的に収束点を探索する。

収束点探索における探索方向ベクトルの決定方法を以下にまとめる。

- 最急降下法; 目的関数の勾配ベクトルの逆方向に探索方向ベクトルを取る。この方法は実装が簡単であるが、収束が遅いと言われている。
- ニュートン法; 目的関数を局所的に2次近似して、探索方向ベクトルを決める。少ない反復回数で収束する。探索方向を決めるためにヘッセ行列 **H** の計算が必要。ヘッセ行列 **H** が正定値であることが前提。レーベンバーグ・マーカート法(修正ニュートン法);ヘッセ行列 **H** が正定値になるように修正をして探索方向を決める。

- 準ニュートン法;ニュートン法に必要なヘッセ行列  $\mathbf{H}$  を近似で行う。
- 共役勾配法;共役勾配方向に探索方向を取る。目的関数が 2 次関数の場合は実装が容易で最終降下法よりも収束が速い。
  - \* 反復法による解法は、「探索ベクトル」の決定方法と「ステップ幅」の決定方法で大きくわけることができる。
  - \* ヘッセ行列  $\mathbf{H}$  は 2 階微分に相当する。なおヘッセ行列  $\mathbf{H}$  が正定値になるとは  $n \times n$  の実対称行列  $\mathbf{H}$  に対して、全ての (0 ベクトルではない)  $n$  次元縦ベクトル  $\vec{x}$  に対して  $\vec{x}^T \mathbf{H} \vec{x}$  が正、もしくは、 $\mathbf{H}$  の固有値が全て正を意味する。

MATLAB 上の関数 `fmincon` では連続して微分が可能な関数であればヘッセ行列  $\mathbf{H}$  を別途定義することでそれを利用した収束点探索が出来る。当初このコマンドを使用して、計算が収束しなかったが、各変数で曲面の傾き `gradient`(一回偏微分)、曲面の向き `hessian`(二階偏微分を書き下した関数行列)を使用することで、最小値探索をすることができた。

距離演算の関数定義において、`[L, grad-vector, hessian-matrix]`の 3 種類の数値を求める MATLAB 関数群を定義する。`fmincon` の中で使用する `@x f(x,A,B,...)`の形で MATLAB 関数宣言し、`optim` セットでその `grad`, `hessian` を使用するを有効にする。`gradient`, `hessian` を定義せずに収束計算を行おうとすると、計算回数の上限を上げても収束に至らないのに対し、`gradient`, `hessian` を定義すると、探索開始点  $\mathbf{x}_0$  を原点やランダムで与えても数分~数 10 分の所要時間で解を得ることができた。より確実に、収束計算回数の上限に至ることがないようにするためには、MATLAB 関数 `lsqlin`, `lsqnonneg` のような範囲制約と線形制約をもつ線形最小二乗法問題のソルバーを用いて探索開始点  $\mathbf{x}_0$  をあらかじめ求めておき、 $\mathbf{x}_0$  を初期値として `fmincon` 関数を利用した。

## 2.5 異なる距離尺度に基づく NMF 法を適用して得た基底ベクトルを用い、ターゲット香気を近似した結果

図 2-09(a)-(d)に示したのはデータ行列  $\mathbf{V}$  もしくはその近似行列  $\mathbf{WH}$  の強度分布である。基底ベクトル数は 30 個である。縦方向に MS データを並べ、各  $m/z$  強度を色調で表現している。各精油香気サンプルのスペクトルを列にして横に並べることで図 2-09(a)のデータ行列の強度分布図を得た。

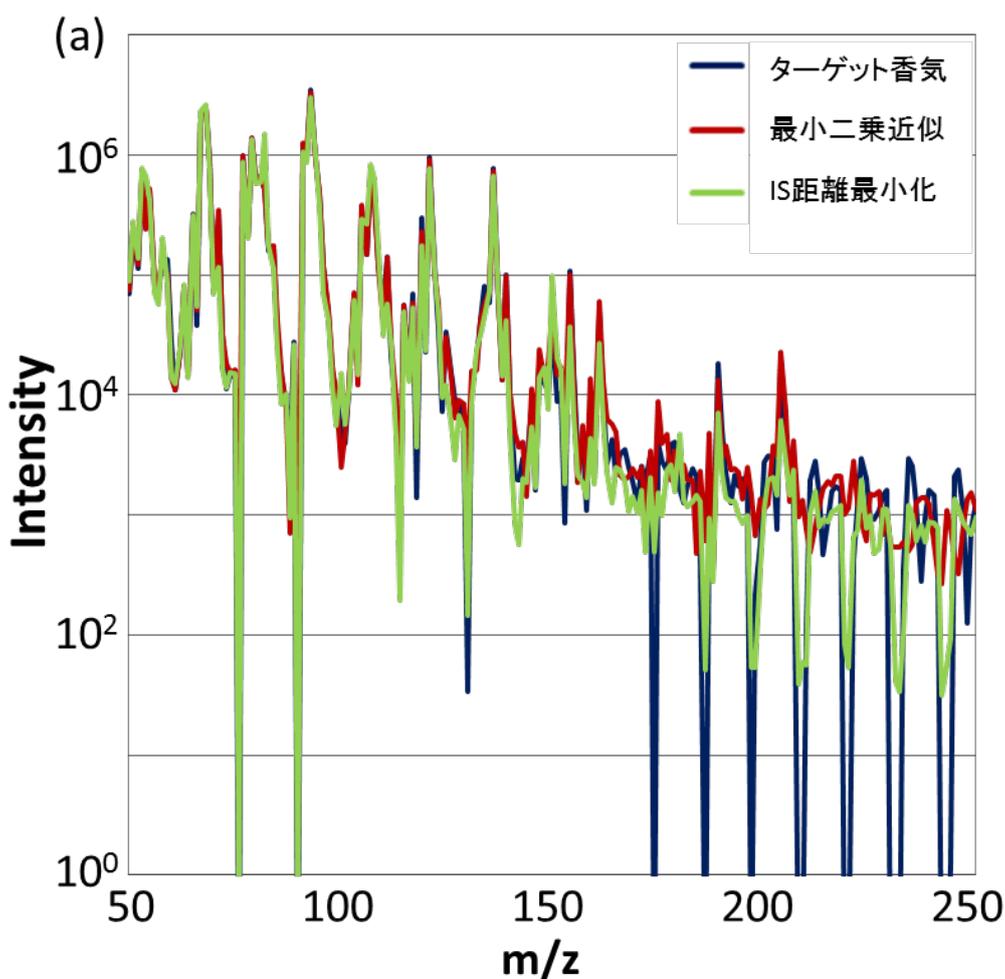
測定データには前述の雑音除去を行っているため、 $\mathbf{V}$  には 0 の要素が現れる。対数でデータを表示するために 0 の値には微小な値(0.1)を加えている。ユークリッド距離、KL ダイバージェンス、IS ダイバージェンスにそれぞれ基づく NMF 法を適用した結果についても、近似行列  $\mathbf{WH}$  の強度を同様に色調で表すと、それぞれ図 2-09(b),(c),(d)になった。同図より図 2-09(b),(c),(d)が高  $m/z$  領域で青で示される弱い強度が多く、元データ(a)図に最も近くなっている。さらに図 2-09(b),(c),(d)の近似行列の全域において、データ行列  $\mathbf{V}$  の要素の値が 0 となっているものについて、 $\mathbf{WH}$  で対応する要素の値が 0 になっているかどうかを、ヒストグラムを作成し出現頻度として確認した。IS 距離の場合、もともとゼロであったデータの再現制度が最も高いことがわかる(図 2-09(b),(c),(d)下段)。



図 2-09(b)よりユークリッド距離に関しては、高  $m/z$  領域のマスペクトルでゼロ点だけでなく、小さなピークにもかなり大きなズレが見られた。それに対して図 2-09(c)の KL ダイバージェンスの高  $m/z$  領域の小さなピークの再現性は高かったが、ゼロ点の再現性が低かった。それに比べると図 2-09(d)の IS ダイバージェンスに関しては高  $m/z$  領域での再現性はゼロ点、小さいピーク共に高かった。

図 2-09(b),(c),(d)との比較から明らかなように、IS ダイバージェンスのゼロ点での再現性は他の 2 種類より優れていることがわかった。高  $m/z$  領域で現れる値は小さいピークや 0 が多いため、ゼロ点の再現性が高いことは高  $m/z$  領域のマスペクトルの再現性を高くすることがここで確認できた。

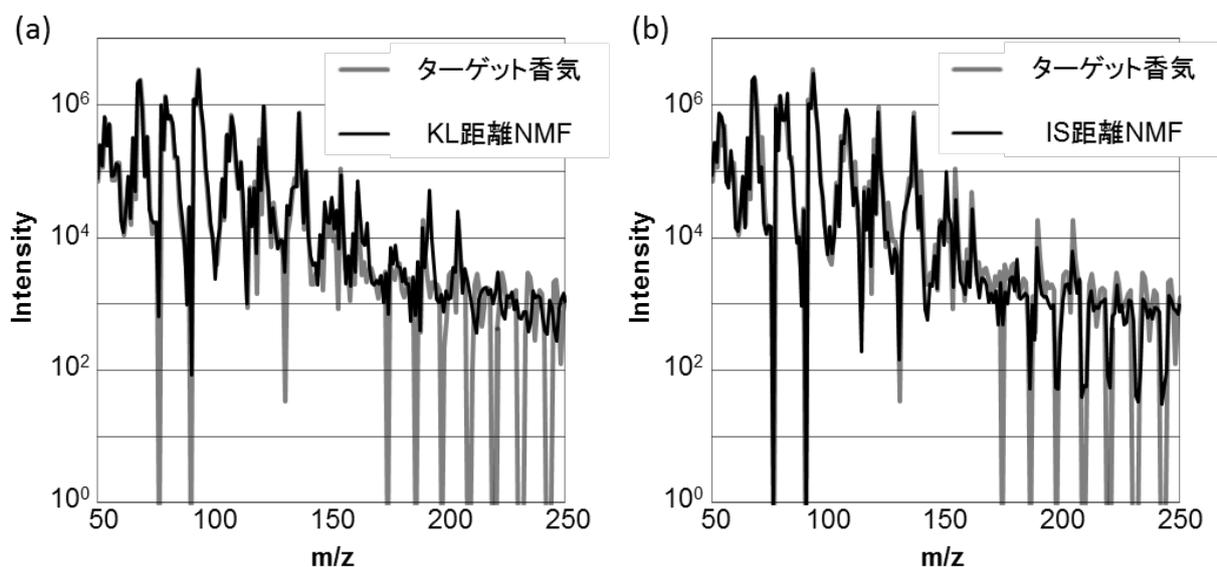
図 2-10 は、IS-NMF を使用した後、Fmincon 関数により IS 距離に基づいて近似臭作成する効果を最小二乗法と比較した結果である。精油データベースに IS 距離を導入した NMF 法の演算を行ったうえで、得られた基底ベクトルを用いた近似臭作成の近似計算において、従来法と IS 距離に基づく方法を比較したものである。どちらの場合もブレンド精油をターゲット香気とし、赤が従来法で近似臭作成に最小二乗法、緑が IS 距離に基づいて近似臭作成をしたスペクトルを示す。



\*いずれも サンプル:精油のみ135、基底数=30,繰り返し回数=10000、ターゲット香気(ブレンド精油Refresh)

図 2-10. 各距離尺度を元にしたターゲット香気の再現 ; (a)近似臭作成に最小二乗法、(b) IS 距離に基づいて、近似臭作成

最小二乗法近似の高  $m/z$  領域の小さなピークの再現性はそれなりに高かったが、ゼロ点の再現性が低かった。IS ダイバージェンスに関しては高  $m/z$  領域での再現性はゼロ点、小さいピーク共に高い再現性が得られることが確認できた。今まで用いられた KL 距離よりも IS 距離に基づく NMF 法では高領域の再現性が高くできることがわかった。再現レシピの導出においても、距離評価に IS 距離を導入することで、高領域の再現性が高められることが分かった。



\* いずれも サンプル; 精油のみ135、基底数=30, 繰り返し回数=10000、ターゲット香気 (ブレンド精油Refresh)

**図 2-11. NMF 計算及び近似計算において各距離尺度を元にしたターゲット香気の再現 ; (a)では NMF 法の距離演算に KL 距離、近似臭作成に最小二乗法を用いた。(b)では IS 距離に基づいて NMF 法の距離演算と近似臭作成を行った**

図 2-11 (a)、(b)は、精油データベースに適用した NMF 法の距離演算と近似臭作成の両近似計算において、従来法 (a)と、IS 距離に基づく方法 (b)を比較したものである。どちらの場合にもターゲット香気であるブレンド精油は共通であり、(a)では NMF 法の距離演算に KL 距離、近似臭作成に最小二乗法を用いた。(b)では IS 距離に基づいて NMF 法の距離演算と近似臭作成をした。

図 2-11(a)に示すように、KL ダイバージェンスの高  $m/z$  領域の小さなピークの再現性はそれなりに高かったが、ゼロ点の再現性が低かった。図 2-11(a)に比べると図 2-11(b)の IS ダイバージェンスに関しては高  $m/z$  領域での再現性はゼロ点、小さいピーク共に高い再現性が得られることが確認できた。

今まで用いられた KL 距離よりも IS 距離に基づく NMF 法では高領域の再現性が高くできることがわかった。NMF 法のみでなく、再現レシピの導出においても、距離評価に IS 距離を導入することで、高領域の再現性が高められることが分かった。

## 2.6 まとめ

精油を用いたマスペクトルデータでは、高領域( $m/z=150\sim 250$ )での強度は低領域( $m/z=50\sim 150$ )に比べて全体的に小さい。しかし、高領域に匂いの特徴が現れやすいため、大きい値と小さい値を同程度に扱うような計算手法が必要であった。基底ベクトルの抽出に IS 距離に基づく NMF 法を用いた結果、IS 距離に関しては高  $m/z$  領域での再現性はゼロ点、小さいピーク共に高かった。IS 距離に基づく最小値探索を導入することで、高領域  $m/z$  のスペクトルの再現性を高められることがわかった。このグラフから、IS ダイバージェンスの方が KL divergence に比べて高領域  $m/z$  の再現性が高いことがわかった。また NMF 法のみでなく、再現レシピの導出においても、距離評価に IS 距離を導入することで、高領域の再現性が高められることが分かった。

IS 距離の発展的な距離尺度として  $\beta$  ダイバージェンスが提案されている。この距離尺度はパラメータを変更することで、連続的にユークリッド距離、KL 距離、IS 距離を定義できる。計算量は多くなってしまいう可能性はあるが、 $\beta$  ダイバージェンスを適用した NMF や近似計算を行うことで、局所解に陥りにくく再現性を高めながら、ダイナミックレンジを適切に設定した要素臭の探索と近似レシピの算出が可能となることが期待される。

### 第三章 嗅覚受容体応答情報に基づいた非負値行列分解法による要素臭の探索

生体の嗅覚に近いセンシング情報を得るためには、より人間の感覚量空間に適したセンシング空間が必要になる点が指摘されてきた。センサアレイを用いて人工的な嗅覚システムの確立を実現するためには、生体の嗅覚のセンシング情報を解析して、人工のセンサアレイ系の応答空間を生体のセンシング空間に合致させることが必要であると考えられる。

**表 3-01. 24 種類のショウジョウバエの ORN 応答を取得した 110 種類の香気物質**

1 ammoniumhydroxide	38 nerol	75 1-octanol
2 putrescine	39 linalool	76 2-pentanol
3 cadaverine	40 $\beta$ -citronellol	77 3-methylbutanol
4 $\gamma$ -butyrolactone	41 linalooloxide	78 3-methyl-2-buten-1-ol
5 $\gamma$ -hexalactone	42 acetaldehyde	79 1-penten-3-ol
6 $\gamma$ -octalactone	43 propanal	80 1-octen-3-ol
7 $\gamma$ -decalactone	44 butanal	81 trans2-hexenol
8 $\delta$ -decalactone	45 pentanal	82 cis2-hexenol
9 methanoicacid	46 hexanal	83 trans3-hexenol
10 aceticacid	47 trans2-hexenal	84 cis3-hexenol
11 propionicacid	48 furfural	85 glycerol
12 butyricacid	49 2-propenal	86 2,3-butanediol
13 pentanoicacid	50 acetone	87 methylacetate
14 hexanoicacid	51 2-butanone	88 ethylacetate
15 heptanoicacid	52 2-pentanone	89 propylacetate
16 octanoicacid	53 2-heptanone	90 butylacetate
17 nonanoicacid	54 6-methyl-5-hepten-2-one	91 pentylacetate
18 linoleicacid	55 2,3-butanedione	92 hexylacetate
19 isobutyricacid	56 phenethylalcohol	93 isobutylacetate
20 isopentanoicacid	57 benzylalcohol	94 isopentylacetate
21 pyruvicacid	58 methylsalicylate	95 trans2-hexenylacetate
22 2-ethylhexanoicacid	59 methylbenzoate	96 methylbutyrate
23 lacticacid	60 ethylbenzoate	97 ethylbutyrate
24 3-methylthio-1-propanol	61 phenethylacetate	98 hexylbutyrate
25 dimethylsulfide	62 benzaldehyde	99 ethyl3-hydroxybutyrate
26 terpinolene	63 phenylacetaldehyde	100 ethylpropionate
27 $\alpha$ -pinene	64 acetophenone	101 ethylmethanoate
28 $\beta$ -pinene	65 ethylcinnamate	102 methylhexanoate
29 (1S)-(+)-3-carene	66 2-methylphenol	103 ethylhexanoate
30 limonene	67 4-ethylguaiaicol	104 hexylhexanoate
31 $\alpha$ -humulene	68 eugenol	105 methyloctanoate
32 $\beta$ -myrcene	69 methanol	106 ethyloctanoate
33 (-)-trans-caryophyllene	70 ethanol	107 ethyldecanoate
34 p-cymene	71 1-propanol	108 ethyltrans-2-butenate
35 geranylacetate	72 1-butanol	109 ethyllactate
36 $\alpha$ -terpineol	73 1-pentanol	110 diethylsuccinate
37 geraniol	74 1-hexanol	

現在、生体の嗅覚情報については幅広い知見が蓄積されつつある。その中でも昆虫については系統的に調べられており、一般的な哺乳類よりも各匂い受容体の機能解明が進んでいる。ショウジョウバエの嗅覚系においては Hallem らが 24 種類の嗅細胞の応答データを報告した[060], [087]。



表 3-02 に示したのは、この応答データである。自然界に一般的に見いだされる表 3-01 に示す 110 種類の香気物質に対するショウジョウバエの 24 種類の ORN の興奮性・抑制性応答が取得された。一定の濃度の 110 種類の匂い物質への各 ORN の応答において、表 3-02 中のマイナスは抑制性応答を表し、プラスは興奮性の応答を表す。興奮性の応答については、プラス数の個数がその強度を表す。“・”は  $n < 50$  spikes/s; “+”は  $50 < n < 100$  spikes/s; “++”は  $100 < n < 150$  spikes/s; “+++”は  $150 < n < 200$  spikes/s; “++++”は  $n > 200$  spikes/s “-”は抑制性の応答を示す)

本章では、生体の嗅覚情報を解析するに当たって、この昆虫の嗅覚応答情報を利用した解析について説明をする。昆虫の嗅覚をモデルとするメリットとしては

- 全嗅覚受容体(嗅覚神経細胞)のデータがそろっている
- 複数の嗅覚神経細胞の信号は脳に送られ脳内で認知が起こる、この認知システムは哺乳類と類似している
- 昆虫匂い受容体は、嗅覚受容体の総数が比較的少なく、機構が単純

という特徴が挙げられる。またこれほど大規模に嗅覚受容細胞について応答を追跡したデータは昆虫以外ではほとんどなく、この Hallem らの 24 種類の ORN の応答データは先駆的な研究である。

### 3.1 NMF 法適用のための応答データの非負化

NMF 法の特徴に、非負のデータから非負の要素を持つ基底ベクトルを抽出する点がある。従来より検討を重ねてきた NMF 法を用いて生体の ORNs の応答を解析するために、興奮性応答と抑制性応答の分離し、応答データの非負化を行った。

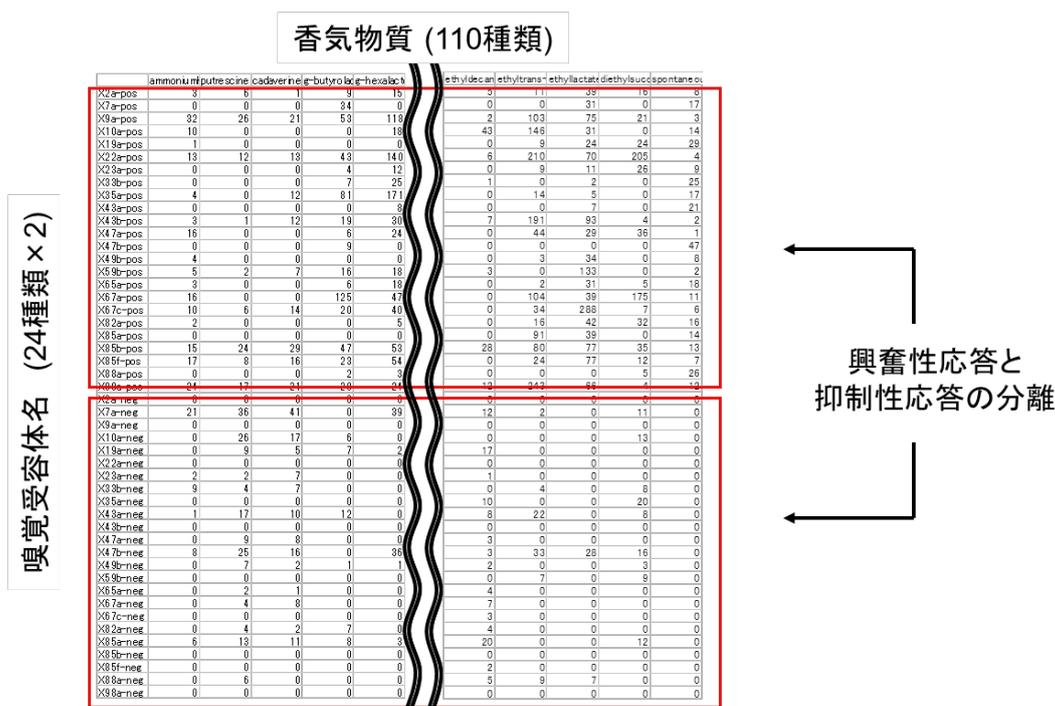


図 3-01. 興奮性応答と抑制性応答の分離と応答データの非負化

今回 Hallem らが報告した Supplemental Data, table S1 [060]をデータとして使用した。このデータでは抑制性の応答が負値で表現されている。NMF 法を用いて、質量分析器データを用いた要素臭の探索の検討と同じような手法で嗅覚受容体反応から基底要素臭を抽出するために、図 3-01 のように興奮性応答と抑制性応答を分離し、抑制性応答を  $\times (-1)$  して別の行で表現することで、非負化した。なお過去の文献例で嗅覚神経細胞の応答レベルを最大1として比例係数で表記し、神経細胞の応答は線形応答であるとしての解析をしている例がある[087]。本検討でも各 ORN に対する最大応答レベルで各応答を割って規格化した。

### 3.2 NMF 法における基底数

まず要素臭の数とデータ行列  $V$  ~ 近似行列  $WH$  の残差の関係について確認した。基底ベクトルの数が増加するほどデータベースの近似精度が向上する反面、基底ベクトル数は必要最小限にしたい。さまざまな基底ベクトル数に対する、残差を求め、プロットし結果を図 3-02 に示す(NMF 条件に関する規格化などの条件は後述)。要素臭の数が増加するほどに残差が改善することが確認できた。一定の繰り返し回数以上で残差減少が止まる事が期待されたが、変曲点は見出されなかった。

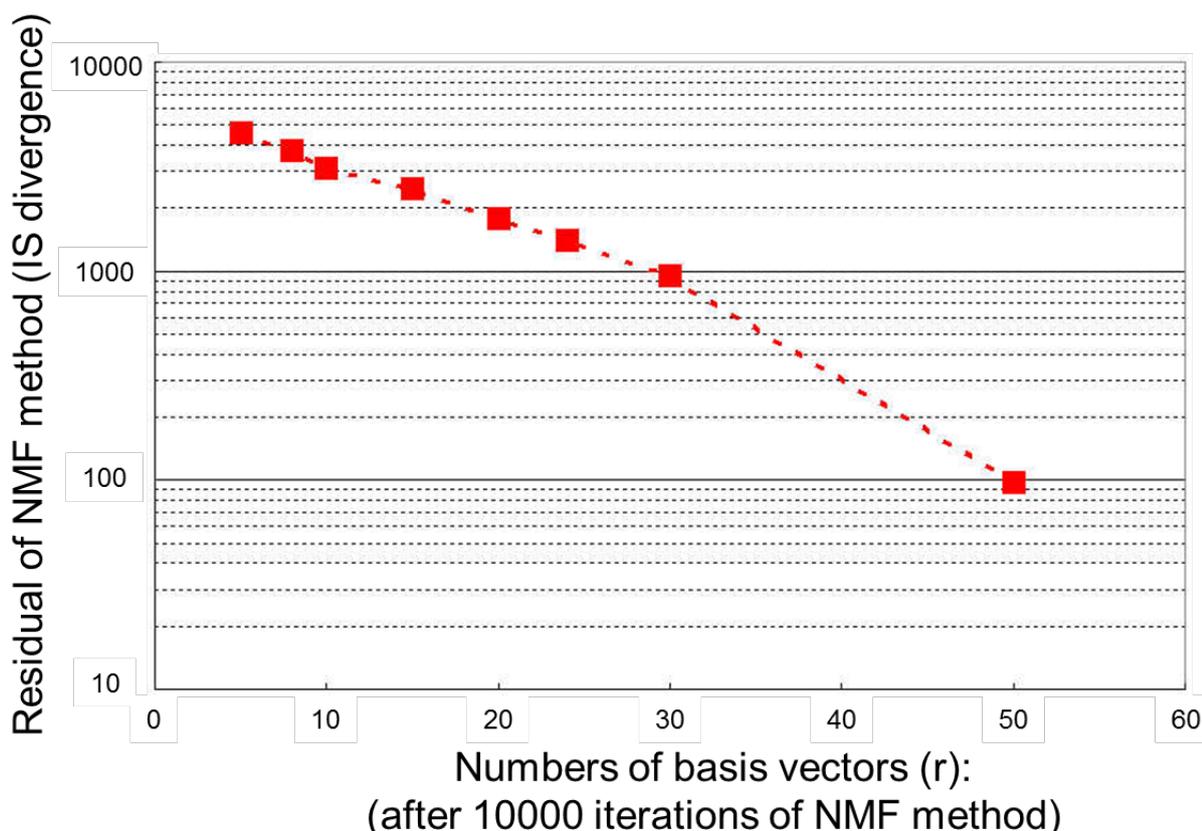


図 3-02. 要素臭の数とデータ  $V$  ~ 近似行列  $WH$  の残差の関係 (各基底数に対し NMF における距離演算は IS ダイバージェンス, 収束計算の繰り返し回数はいずれも 10000 回行い、残差を算出)

本論文においては上記の誤差の検討に加え、元の ORN 個数が 24 個であることを考慮して、基底数を 10 とした場合の検討結果を次節以降にて述べる。

### 3.3 基底行列 $W$ の評価方法と行列因子分解の条件

ここでは基底数  $r=10$ , NMF における距離演算;IS ダイバージェンス、NMF の各行列の規格化条件は表 3-03 に上げた検討条件の二行目(詳細は後述)の条件、収束計算の繰り返し回数はいずれも 10000 回の条件にて NMF 法を実行した結果をもとに、行列因子分解で求められた  $W$  の評価について説明する。

NMF 法で求められた基底行列  $W$  は、元のデータ  $V$  を次元圧縮したものであり、図 3-03 に赤枠で囲った形の基底ベクトルを連結した構造になっている。

NMF 法は乱数に基づく初期行列を繰り返し計算によって基底行列と係数行列を得る。PCA なり、直交性に関わる制約はないので、得られた基底行列の直交性、および複数回実行したときの直交性の再現性を確認した。図 3-03 に合わせて示すのは NMF 法から求められた基底ベクトル間の角度である。白は  $0^\circ$  を表す、色が濃くなるほど  $90^\circ$  に近づく。このように NMF 法によって得られた結果をカラーマッピングで表わすことにより直交性を確認した。NMF 法における  $W$  と  $H$  の初期値を変えて、複数回実行したときの基底行列  $W$  間の直交性の再現性についても同様のカラーマッピングを用いて確認した。

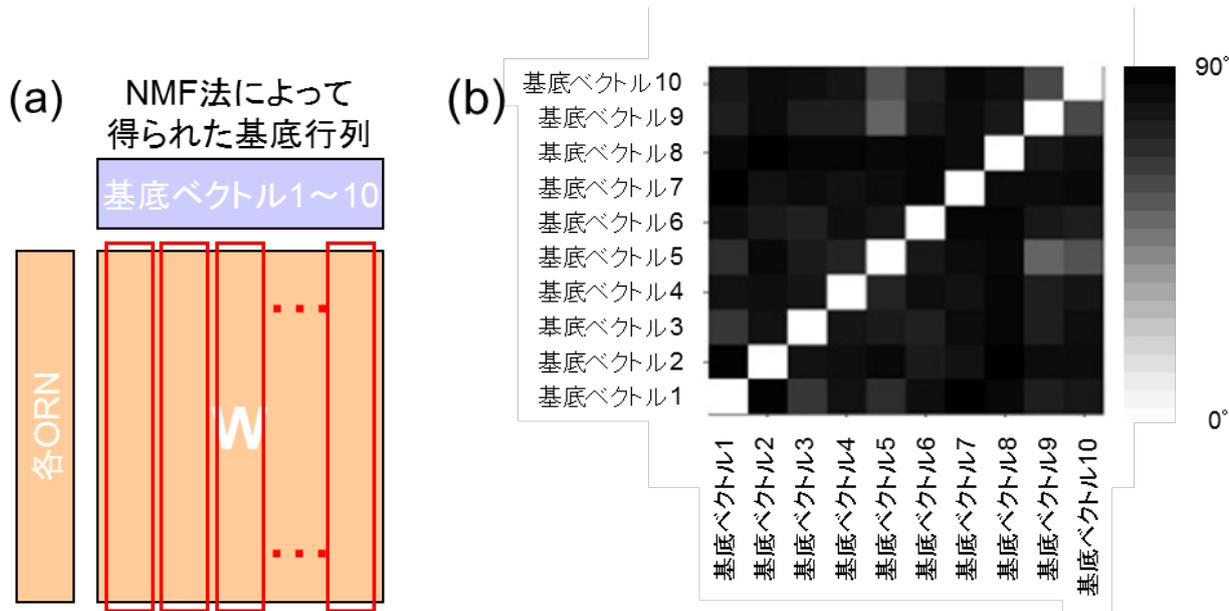


図 3-03. (a) NMF 法で求められた基底ベクトルの直交性確認(白は  $0^\circ$  を表す、色が濃くなるほど  $90^\circ$  に近づく)

基底数  $r=10$ , NMF における距離演算;IS ダイバージェンス, 収束計算の繰り返し回数はいずれも 10000 回

この基底ベクトル間の関係の確認として、各要素のベクトル間にどれくらいの直交性が得られているのかを検証した。その結果、比較的高い直交性がある事が確認できた。またあわせて各ベクトル間の距離を計算し、階層的クラスタリングを行った。各ベクトル間の距離から図 3-04 のような樹状図が得られる。この樹状図では近いものほど近くに存在し、遠いものは遠い。また軸の長さはその距離を反映している。

データ行列  $\mathbf{V}$  の各香気に関して階層的クラスタ解析(距離計算ユークリッド距離)すると、大体香気物質の化合物構造を反映した距離が出る。基底行列  $\mathbf{W}$ 、基底ベクトルを連結した構造になっているので、様々な距離尺度に基づくベクトル間距離が定義できる。図 3-04 に示すのは NMF 法によって得られた基底ベクトルのユークリッド距離による階層的クラスタ分析の結果である。基底行列  $\mathbf{W}$  で得られる各基底ベクトルは仮想的な香気の ORNs によるベクトルになっており元のデータ  $\mathbf{V}$  と合わせて、図 3-06(後述)に示すような 基底ベクトルと初期行列を合わせた階層的クラスタ分析が可能である。

行列因子分解で求められた  $\mathbf{W}$  は、それぞれの化合物クラスターに分散していると考えられる。NMF 法による計算結果は、基底ベクトル間の角度の評価、 $\mathbf{W}+\mathbf{V}$  の階層的クラスタ分析の二点で評価した。

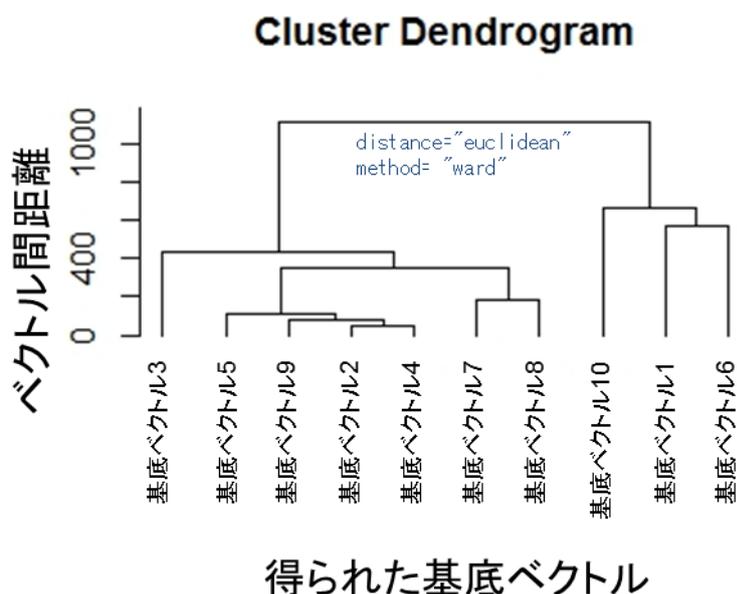


図 3-04. NMF 法によって得られた基底ベクトルの階層的クラスタ分析の結果(V 各基底ベクトルは規格化されていない)

基底数  $r=10$ , NMF における距離演算; IS ダイバージェンス, 収束計算の繰り返し回数はいずれも 10000 回

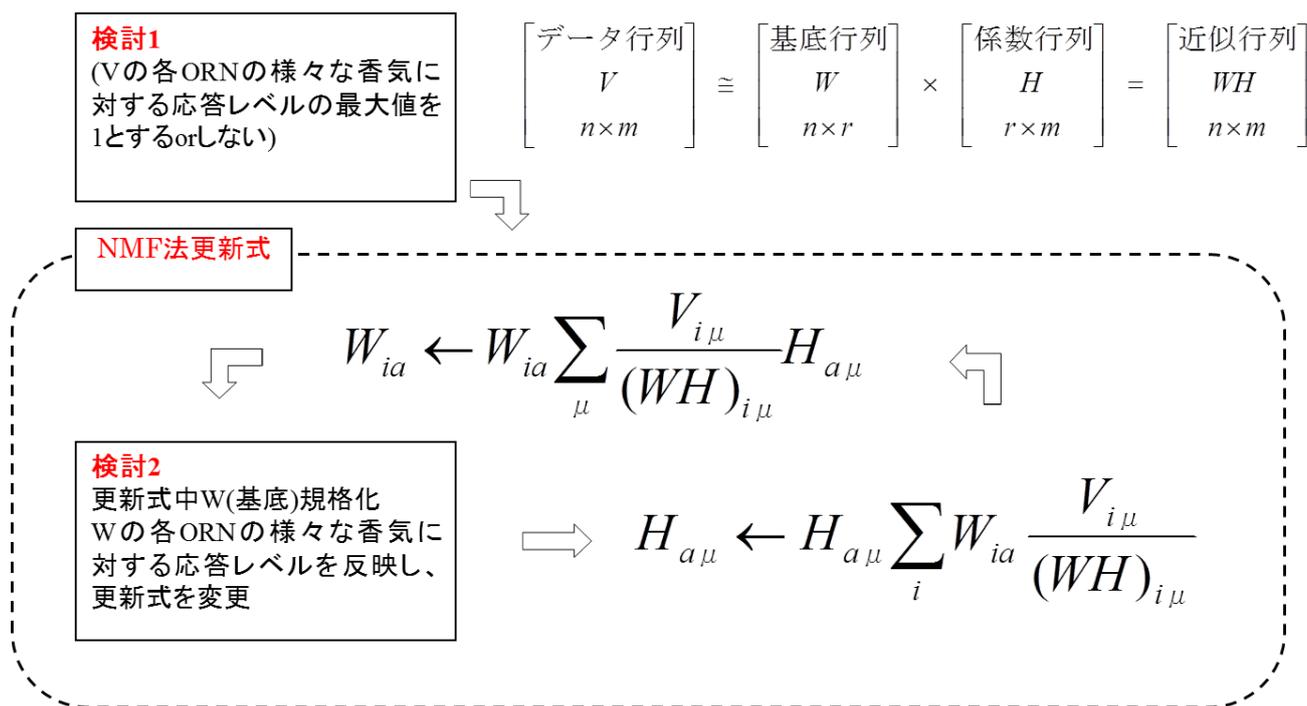


図 3-05. NMF 法において基底行列と係数行列を得る繰返し計算

このような検討において、従来の NMF 法では直交性ではよい結果が出るものの、階層的クラスター解析で確認すると、同じクラスタに複数の基底ベクトルが存在している場合が見いだされた。そこで、図 3-05 に示す検討 1 と検討 2 について、実際の ORN 応答レベルを良く反映した様な応答レベルを導出できるように規格化を変更した。

規格化を変更の検討では、規格化の式を修正して、現実の嗅覚受容細胞の応答レベルを再現するような式に修正した。NMF における距離演算;IS ダイバージェンス, 収束計算の繰り返し回数はいずれも 10000 回である。階層的クラスター分析を使用して、**V** と合わせて **W** を解析し、香気物質のクラスターを代表するような基底ベクトルが得られているかを調査した。得られた基底行列内の直交性、複数回開始値を変えて行った時の直交再現を確認した。NMF 法を適用した結果を用いて Hallem らのデータベースへの近似計算を行った。これにより以下の条件を得た。

- 変更点 1; **V** の規格化を変更。**V** の各 ORN の様々な香気に対する応答レベルの最大値を 1 として規格化
- 変更点 2; **W** の更新式を変更、具体的には **W** の各 ORN の様々な香気に対する応答レベルが **V** と **W** で同等になるようにする。**V** 各行平均=**W** 各行平均になるようにする。

### 3.4 行列因子分解で求められた結果の評価

Hallem らのショウジョウバエの嗅覚応答データを用いた。この応答データはハエの全 24 種類の嗅覚神経細胞の興奮性・抑制性応答から成る。我々は興奮性・抑制性の応答を分離し、応答強度のみとした非負のデータ行列を得た。このデータ行列を階層的クラスター分析すると、香気物質の化学構造を反映した応答を嗅覚神経細胞群がみられることが確認できた。続いてこのデータ行列を NMF 法を適用する際の初期行列とし、NMF 法を適用し

基底ベクトルを算出した。基底ベクトルと初期行列を合わせて階層的クラスタ分析すると図 3-06 に示すような樹状図を得た。

樹状図におけるベクトル間の距離の表され方を確認してみると、例えば左から 2 番目の 1-octen-3-ol と右端から 3 番目のフェニルエチルアルコールは化学構造も大きく異なる物質の組み合わせであるが 24 種類の ORN による信号空間においても大きく距離の離れたものとしてあらわされていることがわかる。対照的にネロールとシトロネロールは両方とも炭素数 10 個の不飽和テルペンアルコールに属し、化学構造もほとんど変わらない物質の組み合わせであり、24 種類の ORN による信号空間においても、基底ベクトル No.7 から右に 4 番目と 5 番目と、距離の離れていないものとして表わされていることがわかる。以上のように、化学物質の構造が異なっているほど距離は長く、化学物質の構造が類似しているほど距離は短くなる傾向が見いだされた。

計算された行列因子分解で求められた各基底ベクトルは図 3-06 中赤枠で囲って示した。信号空間内にそれぞれの化合物クラスターに分散したところにマッピングされており、香気成分のクラスターに合うような基底ベクトルを NMF 法により取り出すことができそうである。

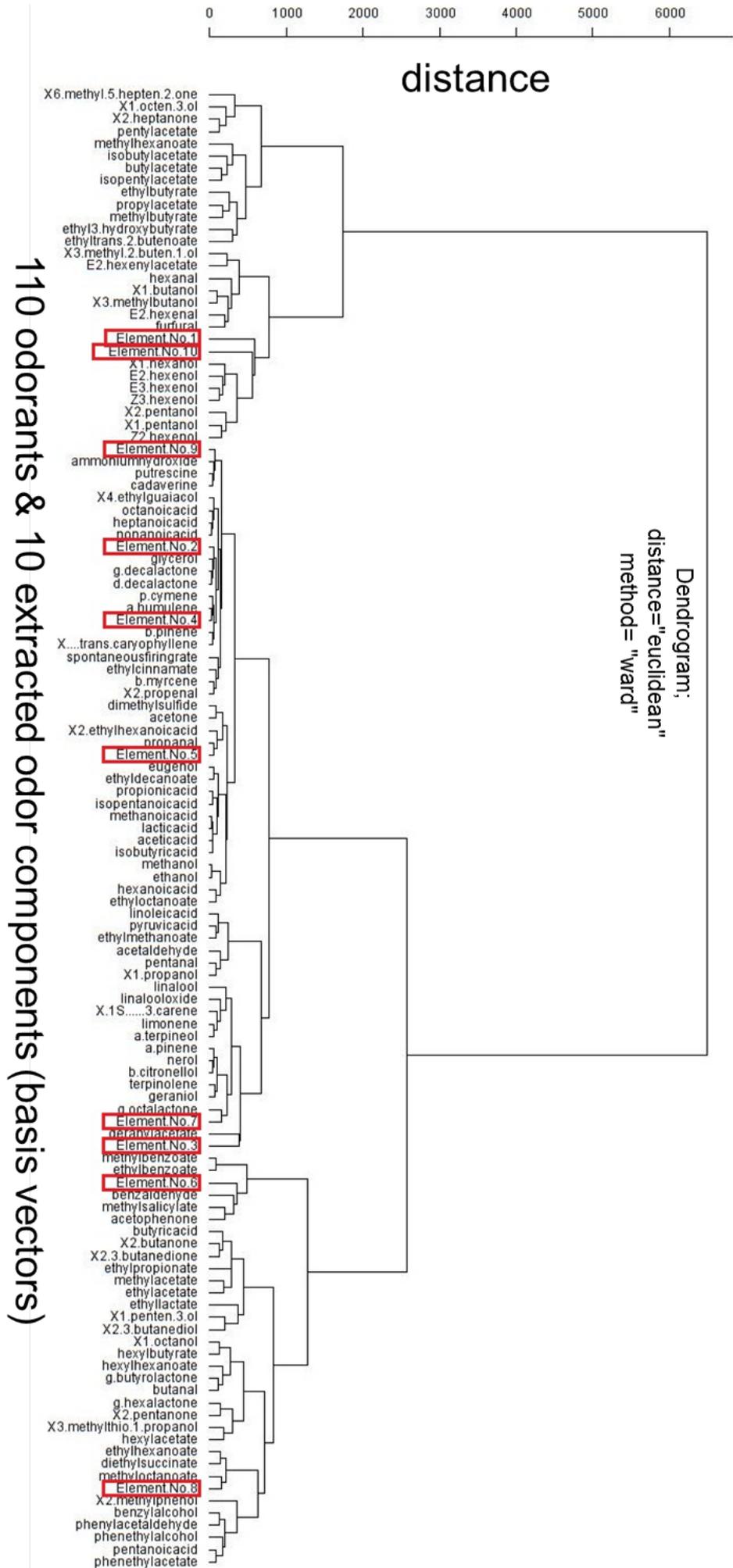


図 3-06. 基底ベクトルと初期行列を合わせた階層的クラスタ分析

### 3.5 10 種類の要素臭を使用した匂い応答パターンの再現

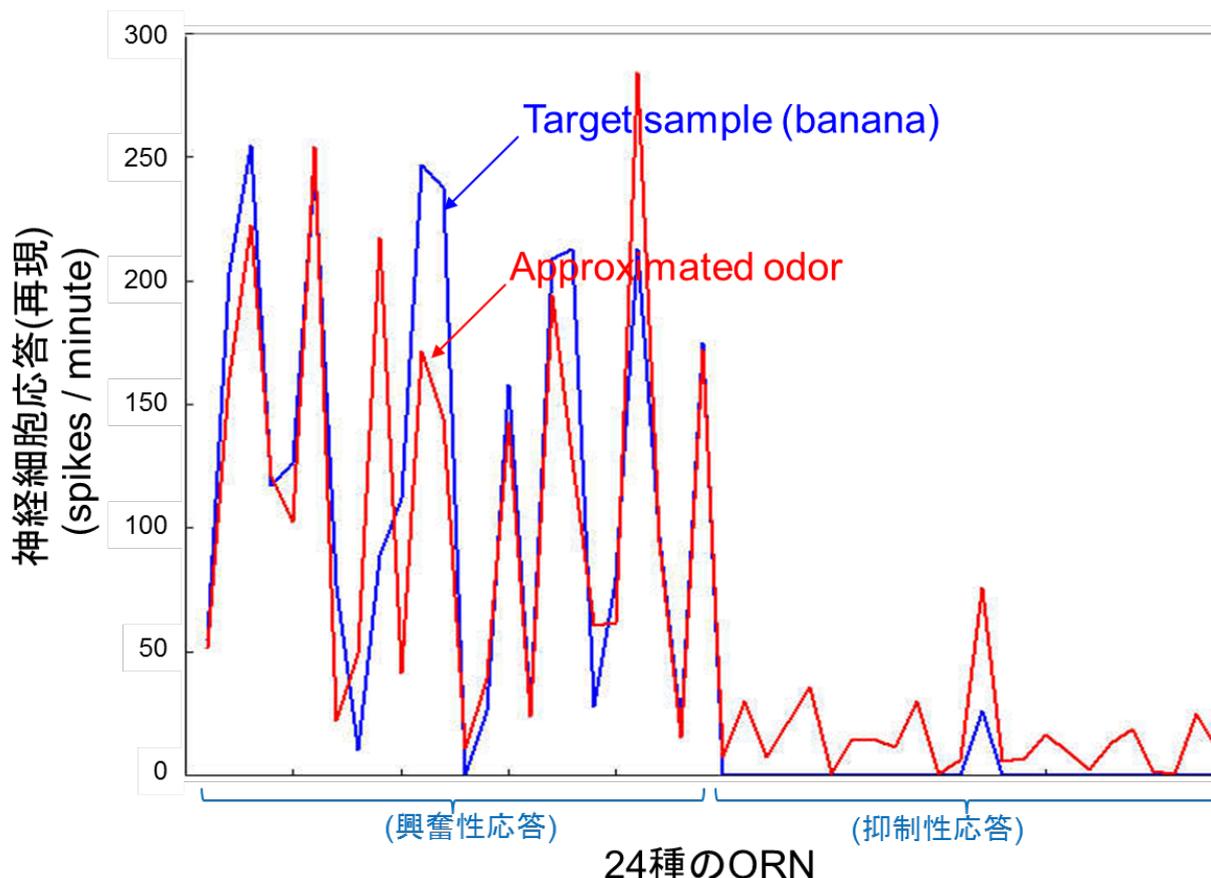


図 3-07. バナナの香りベクトルの再現 ; 青は対象臭を示し、赤は近似臭を示す、応答パターンの相関係数は 0.90

匂いの要素を使用したターゲット香気の匂い応答パターン再現比較を行った。

NMF 法を Hallem らのデータに適用し、基底ベクトルを抽出、得られた基底ベクトルを用いてターゲット香気の ORN 応答を再現する試みた(香りパターンを用いた基底ベクトルの近似計算は行っていない)。ターゲット香気の再現はユークリッド距離に基づく非負拘束付き最小二乗法を用いて行った。非負拘束付き最小二乗法に関しては中本研究室の村上氏、二瓶氏らが報告している。[077]-[079]

今回用いたターゲット香りデータは Hallem らが報告したのものを用いた。本節で説明するバナナ香気は一般のフルーツから有機溶媒を用いて抽出された。香りベクトルはショウジョウバエの全 24 種類の ORN の神経応答を取得したデータである。基底ベクトルを抽出したデータが単一成分の香り物質とは異なり、この香気は混合物に対する全 ORN 応答のものである。得られた基底ベクトルの線形重ね合わせで、その香りベクトルが再現されるとして近似を行った。図 3-07 にバナナの香気の再現した応答パターンを示す。青は対象臭を示し、赤は近似臭を示す、応答パターンの相関係数は 0.90 であった。

以上のように NMF より得られた基底ベクトルを用いて大まかな匂い応答パターンの再現ができた。

### 3.6 NMF 法を用いた要素集作成方法; 匂い近似における要素臭の数と近似臭の精度の相関性の検討

要素臭の数と、要素臭を使用した匂い応答パターンの再現の精度について確認した。基底ベクトルの数が増加するほどにデータベースの近似精度が向上する反面、基底ベクトル数は必要最小限にしたい。いくつかの要素臭の数が異なる基底ベクトルセットを準備し、9種類の果物から抽出した香気のレシピ計算を行った(香りパターンを用いた基底ベクトルの近似計算は行っていない)。基底ベクトル数  $r=5,8,10,15,20,25,30,50$  の条件で、NMFにおける距離演算; IS ダイバージェンス, 収束計算の繰り返し回数はいずれも 10000 回の条件で初期値を変えて NMF 法を 2 回ずつ実行した。各ターゲット香り ORNs 応答と計算された応答再現値の相関係数を算出してプロットした結果が図 3-08 である。オリーブ色の■であらわすプロットが各基底数における平均相関係数である。

9種類の果物から抽出したターゲット香気の匂い応答パターンを基底ベクトルにより再現し比較した結果、要素臭の数が增加するほどにターゲットの香りの近似精度が向上することが確認できた。一定の繰り返し回数以上で残差減少が止まる事が期待されたが、変曲点は見出されなかった。

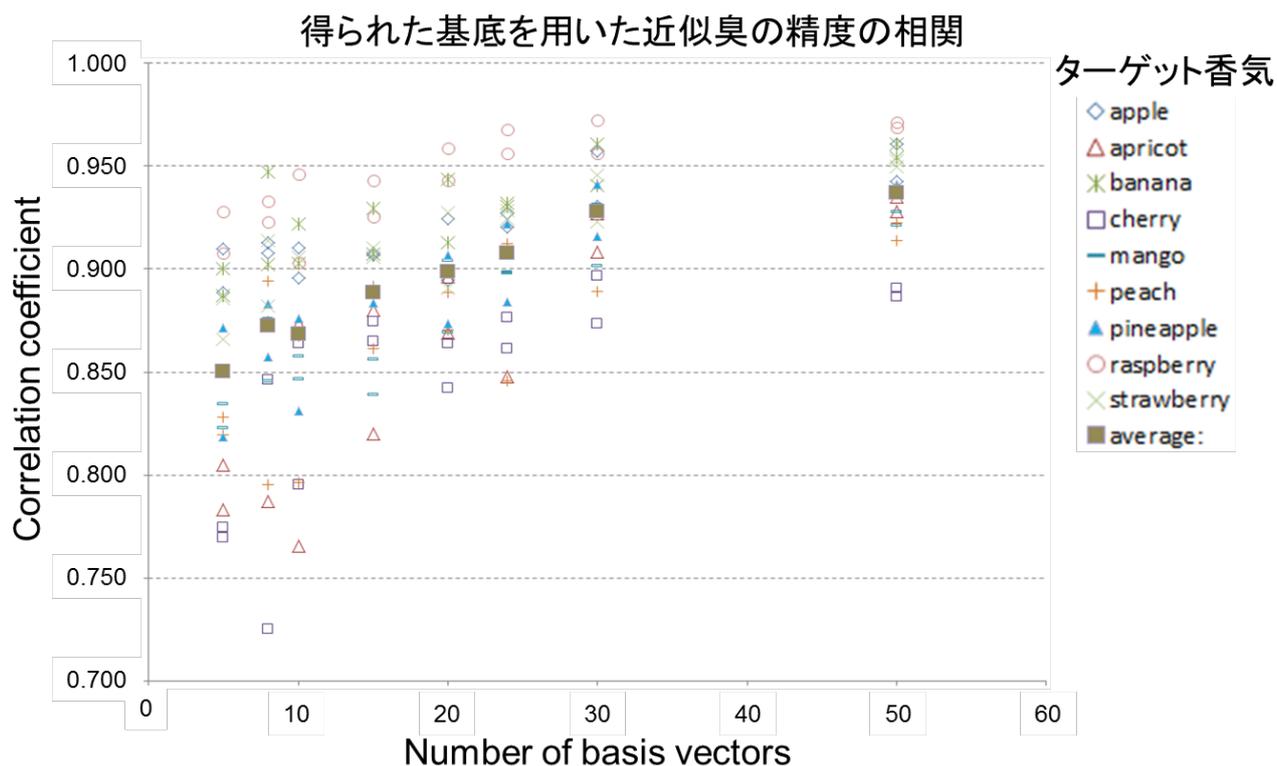


図 3-08. 9種類の天然の果実の香り抽出物に対する ORNs の応答を合成された要素で近似した結果

### 3.7 まとめ

ショウジョウバエの ORN 応答情報の観測結果に NMF 法を適用し、実用的な要素臭の探索について検討をした。NMF 法により抽出された基底を応答調査した香気物質と共に階層的クラスタリングしたところ、抽出された基底ベクトルは、香気成分のクラスタに合うように配置されており、応答が似ている香気分子を代表するものになっていると考えられる。

えられた基底ベクトルの重ねあわせで、現実の香気の ORN 応答が重ね合わせで表現できないか計算した結果、ある程度の相関性を持つ ORN 応答の再現ができる傾向が見られた。また要素臭レシピの計算と得られた計算値上の要素臭を用いて、実際のフルーツから抽出された混合物の香気への ORN 応答を再現した際、計算されたにおい成分はある程度一致した部分がみられた。

だが、昆虫の受容体応答をもとにしたものであり、本来であればヒトの受容体応答をもとにした検討が望まれること、受容体応答が取得された香気物質の種類が 110 種類とかなり制約されたものになっており、匂いの再現が可能になるとは結論できないと考えている。

近似臭作成においては、基底ベクトル数を増大させると NMF における誤差和が減少し、基底ベクトルを用いた近似香気における相関係数が増加することが計算から示された。NMF 法のような数学的手法を用いて実用的な要素臭を探索でき可能性が示唆された。

## 第四章 香気分子構造の数値表現と様々な香気分子によるパラメータ空間

本論文では、人の嗅覚受容体の応答情報は利用できなかった。ヒトの嗅覚受容体の応答情報の取得は一定の成果を上げているが、ORN 全体の機能に関して検討できるほどの詳細な機能はまだ報告されていない。

ここまで筆者は、人間の感覚量空間に適した匂いセンシング空間を、最適な距離尺度を選択した上で、NMF 法などを用いて検討することで、実用的な要素臭の探索に寄与できることを示した。本章では香気分子構造の数値による表現方法と様々な香気分子によるパラメータ空間の解析例に関して述べる

### 4.1 嗅覚受容体の応答強度予測の意義

生物学的な受容体の機能解明データの蓄積は近年急速に進んでいる反面、匂いを構成するすべての香気物質に対する、すべての嗅覚受容体の応答を取得することは事実上不可能であり、使用可能な香気応答データは限定されている。比較的応答情報が得られている昆虫のデータでも限られた化合物に対する応答データしか得られていないことから、受容体の応答予測が必要と考えた。

機械学習による応答特性の予測により、香気物質に限定されていた嗅覚受容空間を図 4-01 に示すように連続的に解析できる。



図 4-01. 香気物質に限定されていた嗅覚受容空間の連続的な解析

筆者は自己組織化写像(SOM)を用いて、嗅覚受容体の応答強度を予測する手法を検討した。NMF 法を用いた解析では、香気物質を ID としてしか扱っていなかったが、筆者はまず化学構造式から各分子を数値パラメータで表現する情報に置き換えた、香気分子の分子構造パラメータ化する手法を探索した。機械学習を含む非線形の予測手法により香気化合物の構造を反映したパラメータをもとに匂いセンシング空間を補間するために課題となることは、数値データの入手方法と選定である。分子の構造を数値表現するものとしては、例えば分子量、元素構成比、官能基(種類ごとの個数)、環状構造(種類ごとの個数)などがあげられる。将来的には結合長・原子位置情報・分子間位置情報を利用することが期待され、さらには分子内の原子位置情報をもとにした電子状態の計算やエネルギー準位の計算を行い、匂いセンシング空間を補間することが期待される。

#### 4.2 パラメータ空間と SOM について

SOM はニューラルネットの一種で、多次元データを圧縮して低次元のマップを得るもので、教師なし学習を適用した解析手法の一種とも言え、SOM は情報構造の可視化のツールとなる。非線形のマッピングにより主成分分析、クラスター分析と同様にデータ間関係の解釈ができる。

これまでに当研究室で SOM を検討の中で使用した人は児玉氏、二瓶氏が SOM を検討している。[077]-[079]

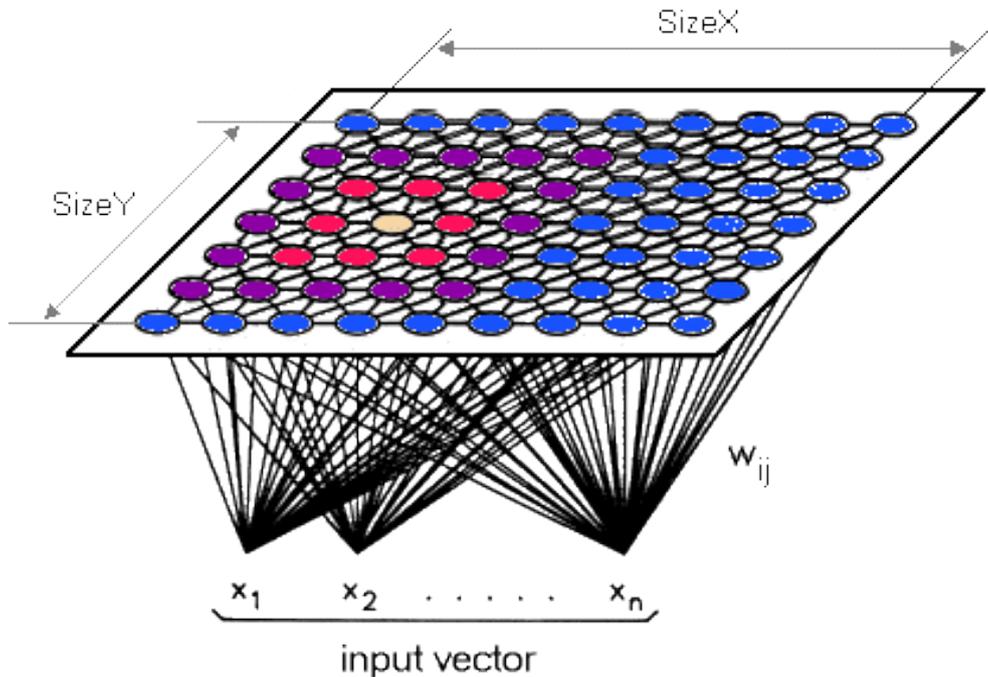


図 4-02. 自己組織化写像(SOM: Self-Organizing Map)

自己組織化写像(SOM: Self-Organizing Map)は、コホネン(T. Kohonen) [092]により提案された教師なしのニューラルネットワークアルゴリズムで、高次元データを低次元空間(主に 2 次元平面上)へ非線形写像するデータ解析方法である。自己組織化写像は、入力層と出力層により構成された 2 層のニューラルネットワークである。出力層は競合層とも呼ばれている。

図 4-02 に 2 次元平面上へ非線形写像する自己組織化写像の基本構造を示す。入力層には分析対象となる  $n$  個の特徴量  $x_1, x_2, \dots, x_n$  からなる特徴ベクトル、出力層には  $X \times Y$  個のユニットがあるとする。図 4-02 で示すように、出力層における任意の 1 つのユニットは、入力層における特徴ベクトルのすべての変数とリンクしている。初期段階では乱数により各変数との間に図 4-02 に示すように参照ベクトル  $W_{ij}$  が付けられている。

SOM のアルゴリズム:

- (1) 入力 と出力層におけるすべてのユニットの中から、最も類似しているユニットを探し出し、そのユニットを勝者とする。
- (2) 勝者のユニットおよびその近傍のユニットの参照ベクトル  $W_{ij}$  を更新する。
- (3) すべての入力特徴ベクトル ( $j=1, 2, \dots, m$ ) に対して(1)~(2)を繰り返す。

SOM に入力ベクトルが与えられると入力ベクトルに最も近い参照ベクトルを持つユニットが勝者ユニットとなる。このときマップ上で勝者ユニットの近くに位置するユニットほど入力ベクトルに対して強く学習する権利を獲得し、その強さに応じて参照ベクトルを入力ベクトルへと近づけるように学習する。このときの学習の強さは勝者ユニットが一番強く、マップ上で位置が勝者ユニットから離れていくほど弱くなる。このような参照ベクトルの更新を色々な入力ベクトルに対して繰り返していくと、ベクトル空間上での入力ベクトル同士の関係性を保存したマップを得ることができる。

SOM は教師なし学習ニューラルネットの一種で、多次元データを圧縮して低次元のマップを描くものである。非線形の主成分分析としての利用や、特に二次元のものは SOM は情報構造の可視化のツールとなり、クラスター分析のような教師なし学習を適用した解析手法ともいえる。

### 4.3 香気物質の分子構造空間の解析

#### 4.3.1 MATLAB における基本的な SOM の実装

```
% matlab 2013b
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%      SOM (Neural Network toolbox 使用)      %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%単層の SOM の検討_動作検証中

%source = 'fisheriris'
load fisheriris          % fisheriris データを読み込み
Label = species;        %トレーニングデータのラベル読込
P = meas;                %トレーニングデータの読込

net = selforgmap([8 8]); %8×8 の大きさの自己組織化写像(未学習)を用意する
net = train(net,P);     %P を用いて自己組織化写像を学習する
view(net)               %自己組織化写像のノード間の距離を着色して図として表示する
y = net(x);             %ベクトルxを学習済みの自己組織化写像に入力すると変換されたベクトルyを得る
y = net(P);             %行列 P を学習済みの自己組織化写像に入力すると変換された行列yを得る

classes = vec2ind(y)    %ベクトルをユニットのインデックスに変換
% ここまで
```

**図 4-03. コマンド例**

neural network toolbox で提供されるパッケージ化された SOM コマンド群がある。これを用いて、得られた SOM を構築し、未知の信号を分類または未知の信号から得られる固有値の推測ができる。構築された SOM の元の教師信号がどこに投射されているのか、を決定するときにも neural network toolbox で提供されるパッケージ化された SOM コマンドでおこなう。たとえば図 4-03 に iris のデータをもとに 8×8 の SOM をトレーニングするコマンド例を示す。データを用いて自己組織化写像を学習する課程を単純なコマンドで処理することができる。図 4-04 に neural network toolbox で提供される SOM の例を示した。図 4-04(A)のように SOM のレイヤーのユニットの配列として'hextop'(平面 hexagonal 状=ハニカム状(六辺形)にユニットを並べる)トポロジーを使用した、紺色の六角形で示される各ノード間には赤い線で示されるノード間距離が定義できる。学習が終わるとノード間距離が計算でき、

コマンド `view(net)` を使用すると、図 4-04 (B) のように自己組織化写像のノード間の距離を着色して、図として表示できる。ノード間距離は近い方から黄色→オレンジ→赤→黒で示される。これに加えてトレーニングデータを各ノードにマッピングしながら、検討と考察を行った。

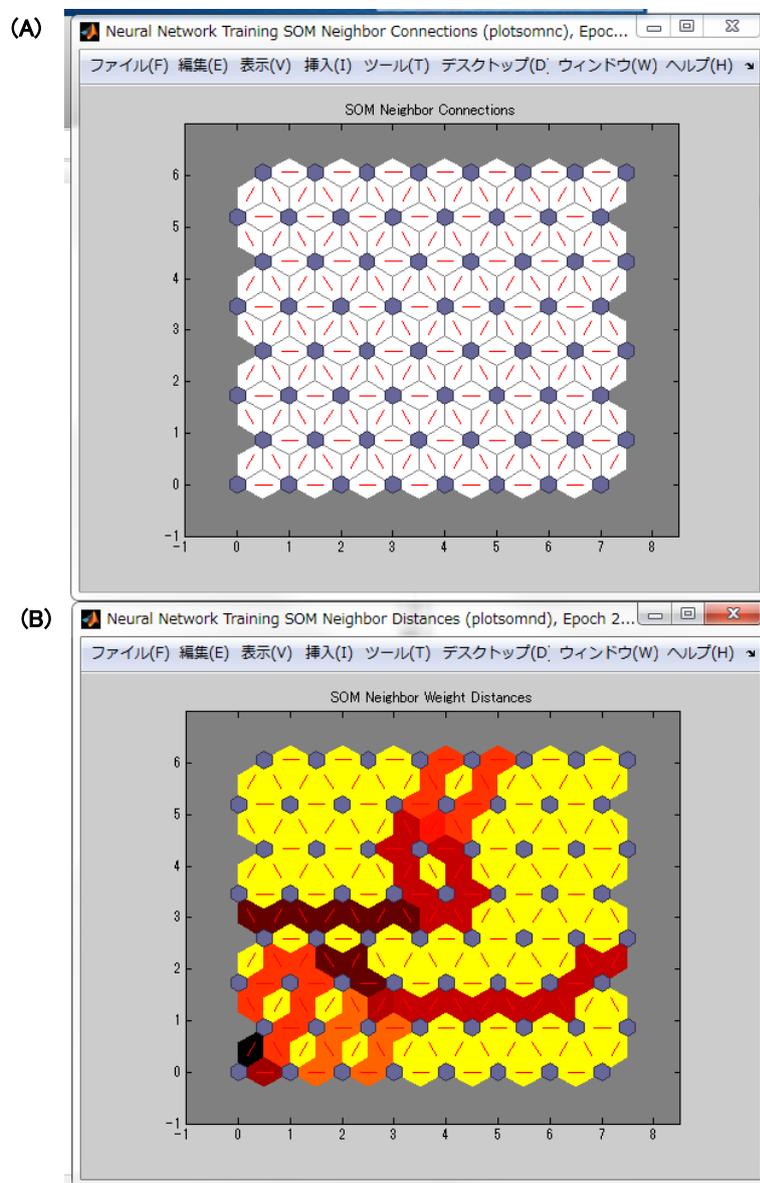


図 4-04. neural network toolbox で提供される SOM の例 ; (A) SOM 学習前 `view(net)` の出力、(B) 学習後の `view(net)` のノード間の距離を色調表示した出力(テストサンプル ; fisher iris データ)

#### 4.4 化合物の構造表記方法と分子構造パラメータ

##### 4.4.1 コンピューター上で取り扱いやすい化合物の構造表記方法

通常の化学においてみられる構造式や組成式から、コンピューター上で取り扱いやすく異性体を正確に区別できる表記法にいったん変換したうえで、香気分子の分子構造パラメータ化した。構造異性体の表記方法ではどこ

まで化合物を区別できるか、という点が重要である。最も完璧な表現方法は化合物の構造表現方法は構造式で表現することであるが、「図」なのでコンピューター上では取り扱えない。化学物質構造をデータとして取り扱う方法としては以下のようなものがある。

- 化合物名(物質を正確に特定できる方法); IUPAC 命名法(図 4-05a), CAS No.(図 4-05b),物質情報の管理、検索やネット上の情報収集に便利である
- 構造を反映した表記(線形表記方法); SMILES(図 4-05c)、Finger Print 法(MDL Maccs key、Daylight FP、ECFP6)
- 構造を反映した表記; 構造式(図 4-05d 人間にとって最もわかりやすく、細かい情報も与えてくれる表記方法) …ただし「図」なのでコンピューター上では取り扱えない
- 構造を反映した表記(行列表記法); MDL-MOL(図 4-05e)、SDF(MOL の変法)…行列は拡張され、原子位置座標も含めた表記もともなっているのが通常

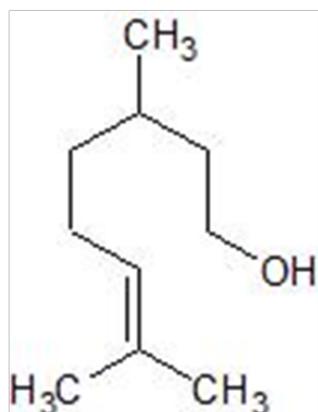
Citronellol;

(a) IUPAC Name;  
3,7-dimethyloct-6-en-1-ol

(b) CAS;  
106-22-9

(c) SMILES;  
CC(CCC=C(C)C)CCO

(d) 構造式;



(e) MDL-mol;

```

C10H20O
MOLSOFT 01201615292D

11 10 0 0 0 0 0 0 0 0999 V2000
1.4000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
2.1000 1.2124 0.0000 C 0 0 3 0 0 0 0 0 0 0 0 0
3.5000 1.2124 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
4.2000 2.4249 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.6000 2.4249 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
6.3000 3.6373 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.6000 4.8497 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
7.7000 3.6373 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
1.4000 2.4249 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
-0.0000 2.4249 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
-0.7000 3.6373 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
2 1 1 0 0 0 0
3 4 1 0 0 0 0
4 5 1 0 0 0 0
5 6 2 0 0 0 0
6 7 1 0 0 0 0
6 8 1 0 0 0 0
2 3 1 0 0 0 0
2 9 1 0 0 0 0
9 10 1 0 0 0 0
10 11 1 0 0 0 0
M END
$$$$

```

図 4-05. 様々な化合物の構造表記方法(例: シトロネロール)

分子の形状をデータベースや化学計算で利用可能にするための手法は大別すると2種類に分けられる。線形表記方法は1バイト文字の鎖状で表現する手法であり、行列表記法は1バイト文字の鎖状で含有する原子を表し、かつ分子内の結合を数値行列で表現してあるので、位置情報を含む量子化学計算にも対応可能な表記方法である。[093]

筆者は本検討で、線形の表記法として広く知られていてデータベース上で扱いやすい SMILES を使用した。

なお、分子構造表記への変換によって得られるデータは、化合物の構造情報を利用した分子認識空間の解析に直接利用できるわけではない。東大嘉手川氏(神崎研)も以下の手順で 5000 種類の分子構造パラメータを得、化合物の構造情報を利用した分子認識空間の解析に利用した。

- ①化合物名から分子構造表記(MDL-mol 式)へ変換
- ②分子構造からプログラムを用いた分子構造のパラメータを抽出

構造式や組成式から、コンピューター上で取り扱いやすく異性体を正確に区別できる表記法にいったん変換したうえで、香気分子の分子構造パラメータ化する作業を効率的に行うためには、自動化を検討したほうが良い。筆者は pubchem(=pubmed)について調査したが、使用方法、自動的にスクリプトで情報を引いてくる手法が不明瞭であったので、別の手法を探した。pubchem 以外では、データベースから python 上で動作する化学情報サイト chemspider [094]の API へのアクセスを可能にする Python モジュール 'cirpy'を追加することで、化合物名から CAS No.をリストアップし、化合物構造表記の出力が可能である。

#### 4.4.2 嗅覚受容体の応答強度予測に用いる分子構造情報

嘉手川氏(神崎研)は、それぞれの化合物から約 5000 種類の分子情報を取得したうえで、110 種類の香気物質に対する応答から各受容体の受容応答予測における分子情報の取捨を検討した[095]。嗅覚受容体の分子認識機構において、どのような分子構造が影響を与えているのかを明らかにするが、分子生物学では期待されている。嘉手川氏(神崎研)の研究においては、分子情報には分子構造情報のみからなり、沸点や溶解性などの物性情報は入っていないようである。

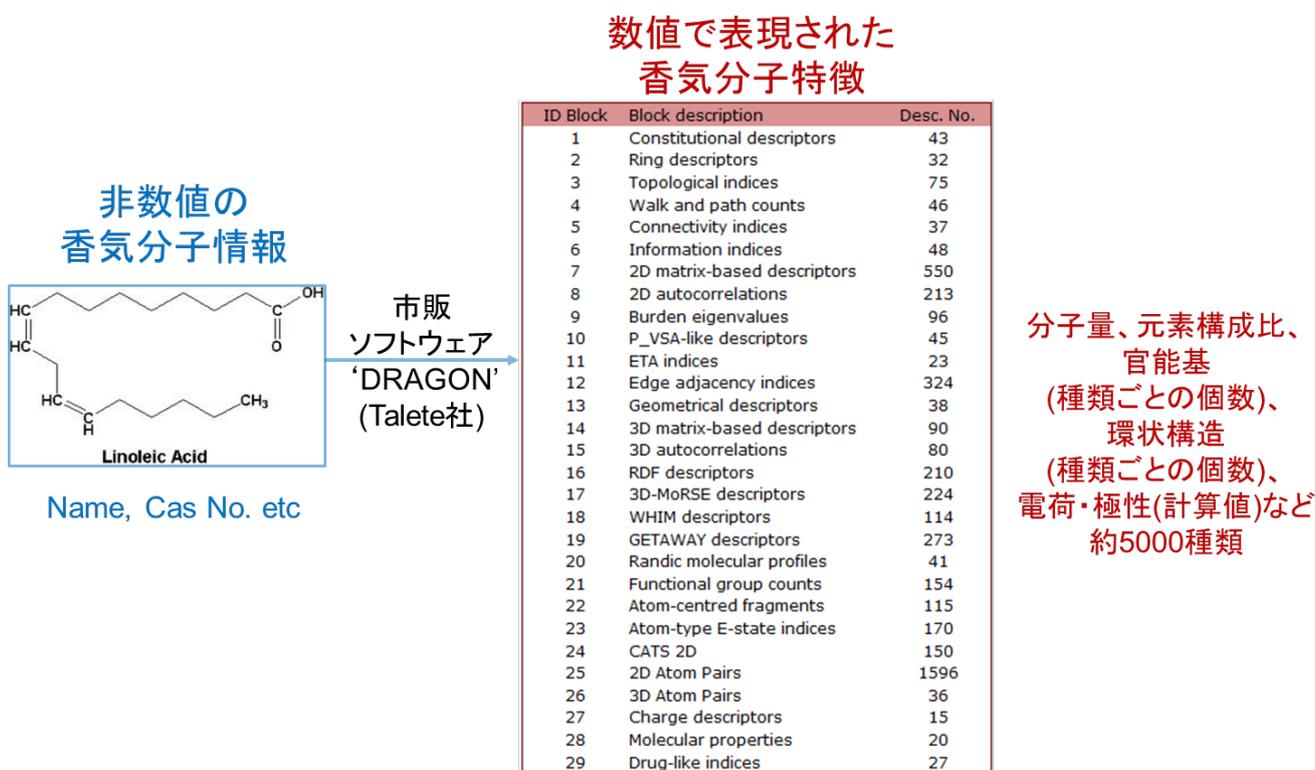


図 4-06. Talete DRAGON により得られる個々の香気分子に対応する分子構造パラメータ

Talete DRAGON(最新版 6.0)は日本国内代理店 Affinity Science Corporation が提供するソフトウェア料金¥100万円前後のソフトウェアである。分子記述子は、分子の構造活性相関や構造物性相関、類似性/多様性解析、分子データベースのスクリーニング(HTS)などに利用することができ、新薬開発の活性相関等の検証に使用することを想定されている。図 4-06 に示すようにそれぞれの化合物から約 5000 種類の分子情報を取得することができる。

このソフトへの構造式入力には MDL-MOL、および SMILES でも入力が可能で、テキスト形式で SMILES を読み込ませることもできる。嘉手川氏(神崎研)の修論では、生物系のデータベース pubchem を利用し MDL-MOL 分子構造表記に変換したうえで、Talete DRAGON への入力を行い、5000 種類の分子パラメータを抽出していた。本研究では神崎研究室により、Talete DRAGON を使用して取得した全香気分子の分子構造情報のテーブル(約 110 種類の香気物質×約 5000 種類の分子構造パラメータ(形式は xls)を提供していただいた。

嘉手川氏(神崎研)の分子パラメータと比較して、DRAGON から NaN が返されたパラメータの個数、ゼロが返されたパラメータの個数が異なっていた。データの解析においては本章の内容を嘉手川氏(神崎研)のパラメータ化、次章以降の検討を筆者のパラメータ化に基づき検討した。それは構造式が同じでも、CAS No.の違いや、MDL-MOL、および SMILES といったソフトへの構造式入力の違いによるものと考えている。

最新の DRAGON では 4882 種類のパラメータが抽出できる反面、嘉手川氏(神崎研)が用いたパラメータは 4868 種類と、幾つかのパラメータが欠けた状態である。この状態で嘉手川氏(神崎研)が応答解析を行っている。その欠如したパラメータは現行の Talete DRAGON におけるパラメータの番号;138, 145, 4824-4838(連番)→計 17 個であった。これらのパラメータ名を表 4-01 にあらかわす。今回検討において、分子構造パラメータが若干異なっているが大勢には影響なしと考えて検討を行った。

**表 4-01. 嘉手川氏(神崎研)が応答解析に使用しなかった Talete DRAGON の分子構造パラメータ**

パラメータの番号	パラメータ名
134	Psi_i_s
145	Psi_e_t
4824	qpmax
4825	qnmax
4826	Qpos
4827	Qneg
4828	Qtot
4829	Qmean
4830	Q2
4831	RPCG
4832	RNCG
4833	SPP
4834	TE1
4835	TE2
4836	PCWTE1
4837	PCWTE2
4838	LDI

## 4.5 パラメータ空間と SOM

### 4.5.1 香気物質の空間に基づく、パラメータ空間の非線形マッピング

Hallemらの香気物質に関する Tallete DRAGON によるパラメータの分散について、データ構造の解析を示す。まず変動していないパラメータや NaN の形でエラーが出ていないデータのみで成型した。パラメータ数は 2831 個になった。SOM の学習回数は 100 回、初期近傍サイズは 3、SOM のレイヤーのユニットの配列のトポロジーは 'hextop'(平面 hexagonal 状=ハニカム状(六辺形)にユニットを並べる)、ニューロンの距離関数は 'linkdist'(=link distance)とし、そのほかのパラメータは neural network toolbox パッケージで提供される SOM コマンドのデフォルト値を用いた。学習パラメータにおいて、SOM サイズのみ検討し(後述)、以降他の検討ではこの条件を用いて検討した。学習後の SOM のノード間距離を近い方から黄色→オレンジ→赤→黒で示すと、ノード間距離は図 4-07 で示される。図 4-07 に嘉手川氏(神崎研)の Tallete DRAGON によるデータを SOM 解析し、110 種類の香気物質をマッピングした結果を示す。数値は化合物の番号を示す。一部のノードについてのみ大きな変化が見られており、大部分については平坦なデータを持っていることが分かった。このことから SOM による非線形写像を作成する際にも規格化条件などを検討する必要性がある。

1 ammoniumhydroxide	38 nerol	75 1-octanol
2 putrescine	39 linalool	76 2-pentanol
3 cadaverine	40 b-citronellol	77 3-methylbutanol
4 g-butylolactone	41 linalooloxide	78 3-methyl-2-buten-1-ol
5 g-hexalactone	42 acetaldehyde	79 1-penten-3-ol
6 g-octalactone	43 propanal	80 1-octen-3-ol
7 g-decalactone	44 butanal	81 E2-hexenol
8 d-decalactone	45 pentanal	82 Z2-hexenol
9 methanoicacid	46 hexanal	83 E3-hexenol
10 aceticacid	47 E2-hexenal	84 Z3-hexenol
11 propionicacid	48 furfural	85 glycerol
12 butyricacid	49 2-propenal	86 2,3-butanediol
13 pentanoicacid	50 acetone	87 methylacetate
14 hexanoicacid	51 2-butanone	88 ethylacetate
15 heptanoicacid	52 2-pentanone	89 propylacetate
16 octanoicacid	53 2-heptanone	90 butylacetate
17 nonanoicacid	54 6-methyl-5-hepten-2-one	91 pentylacetate
18 linoleicacid	55 2,3-butanediol	92 hexylacetate
19 isobutyricacid	56 phenethylalcohol	93 isobutylacetate
20 isopentanoicacid	57 benzylalcohol	94 isopentylacetate
21 pyruvicacid	58 methylsalicylate	95 E2-hexenylacetate
22 2-ethylhexanoicacid	59 methylbenzoate	96 methylbutyrate
23 lacticacid	60 ethylbenzoate	97 ethylbutyrate
24 3-methylthio-1-propanol	61 phenethylacetate	98 hexylbutyrate
25 dimethylsulfide	62 benzaldehyde	99 ethyl3-hydroxybutyrate
26 terpinolene	63 phenylacetaldehyde	100 ethylpropionate
27 a-pinene	64 acetophenone	101 ethylmethanoate
28 b-pinene	65 ethylcinnamate	102 methylhexanoate
29 (1S)(+)-3-carene	66 2-methylphenol	103 ethylhexanoate
30 limonene	67 4-ethylguaiacol	104 hexylhexanoate
31 a-humulene	68 eugenol	105 methyloctanoate
32 b-myrcene	69 methanol	106 ethyloctanoate
33 (-)-trans-caryophyllene	70 ethanol	107 ethyldecanoate
34 p-cymene	71 1-propanol	108 ethyltrans-2-butenolate
35 geranylacetate	72 1-butanol	109 ethylacetate
36 a-terpinol	73 1-pentanol	110 diethylsuccinate
37 geraniol	74 1-hexanol	

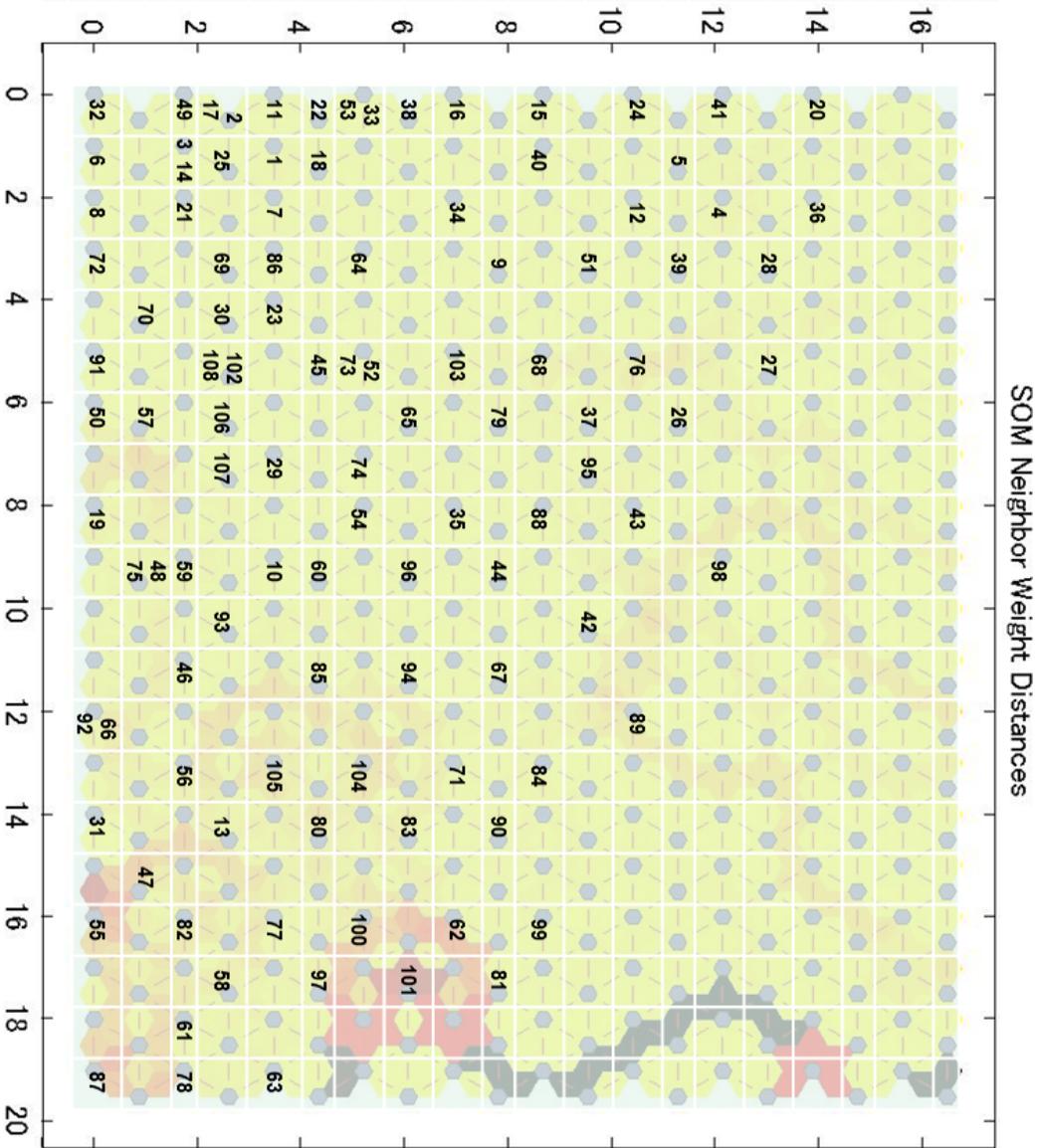


図 4-07. 図分子情報の SOM (Talele DRAGON 導出・サンプル数 110・次元数 2831) 学習回数 200 回、layer 20 x 20

#### 4.5.2 化合物の構造パラメータ空間の非線形マッピング

SOM はニューラルネットの一種で、多次元データを圧縮して低次元のマッピングを得ることができ、情報構造の可視化のツールとなる。図 4-08 に示すのは、市販の約 1500 種類の香料化合物の分子構造をもとに化合物群のマッピングを得る例である。

市場に出回っている香気物質を調査し全般的な香気物質空間に基づいて、使用した香気物質を選択したり、写像の結果を考察する必要がある。筆者は印藤の「合成香料」[088]を参考にして市販の約 1500 種類の香料化合物のリストを作成し、SMILES 表記式へ変換した後を用いて、Talete DRAGON を用いてそれぞれの化合物に対し約 5000 種類の分子構造パラメータを得た。このマッピングへの注目している物質の写像を調査することにより、物質と似ている物質・異なる物質を探ることが可能である。

日本語版の印藤の「合成香料」[088]を用いたが、海外では Steffen Arctander, 「Perfume and Flavor Chemicals」[089]が広く知られているため、合わせての検討が望ましい。



図 4-08 パラメータ空間の非線形マッピング

Talete DRAGON によるパラメータは大部分が平坦なデータを持っていることが前節図 4-07 からわかる。SOM による非線形写像を作成する際の規格化条件などの検討の必要性はあるが、本研究では変動していないパラメータや NaN の形でエラーが出ていないデータのみで成型し、パラメータ毎にオートスケールリング(標準偏差 1 平均 0 になるように規格化)を行い、約 1500 種類の香気物質の番号をマッピングした。結果は図 4-09 に示す。図中において、学習後の SOM のノード間距離を近い方から黄色→オレンジ→赤→黒で示す(番号は香気物質の ID)。

なお筆者は約 1500 種類の香気物質のマッピングを応用して、香気物質の空間に基づく能動的なデータ収集手法を検討した(5 章にて後述)。

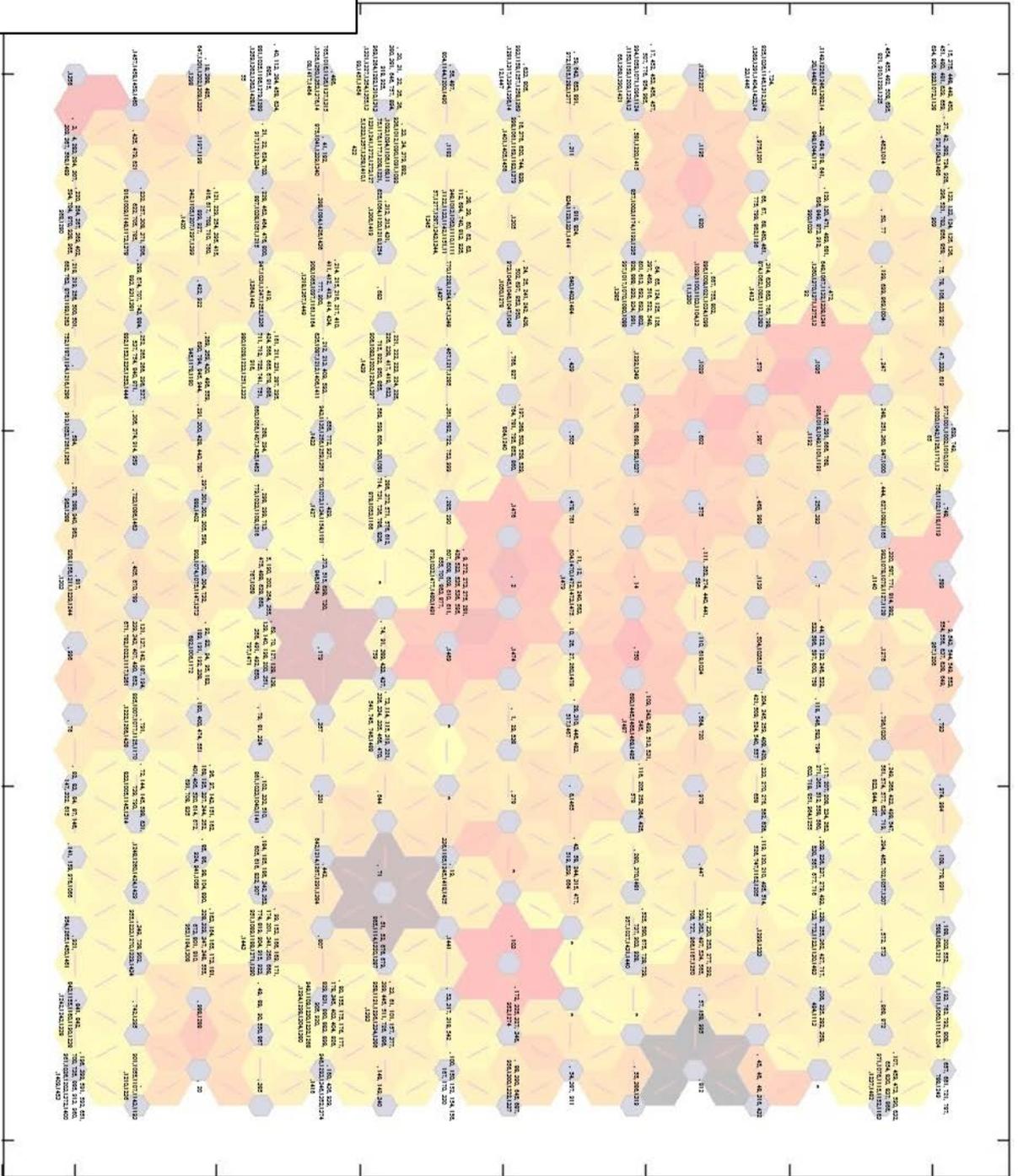
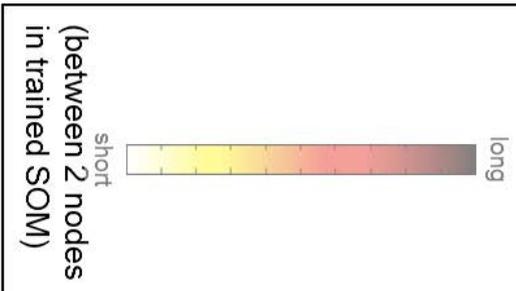


図 4-09.学習後の SOM への約 1500 種類の香気物質のマッピング

#### 4.6 まとめ

匂いを構成するすべての香気物質に対する、すべての嗅覚受容体の応答を取得することは事実上不可能であり、使用可能な香気応答データは限定されている。比較的応答情報が得られている昆虫のデータでも限られた化合物に対する応答データしか得られていないことから、受容体の応答予測が必要と考えた。

化合物の構造表現方法は構造式で表現することが一般的だが、コンピューター上では取り扱えない。本検討では線形の表記法として広く知られていてデータベース上で扱いやすい **SMILES** を使用した。

分子構造を線形表記しただけでは、化合物の構造情報解析は直接利用可能とはならない。今回は約 5000 種類の分子記述子に基づく分子構造パラメータを与えてくれるソフトウェア **Taleta DRAGON** を用いて、香気分子の分子構造を数値表現し、分子構造情報空間が検討可能になることを確認した。

**SOM** はニューラルネットの一種で、多次元データを圧縮して低次元のマッピングを得ることができ、情報構造の可視化のツールとなる。**SOM** を利用して、市販の約 1500 種類の香料化合物の分子構造をもとに化合物の分子構造情報空間のマッピングを得ることができた。

# 第五章 香気分子構造をもとにした自己組織化写像による嗅覚受容体の応答強度予測手法

3章のように ORNs の応答データ[060]をもとに要素臭を抽出し、近似臭を作成できる可能性が示された[081]。生物学的な受容体の機能解明データの蓄積は近年急速に進んでいる反面、匂いを構成する香気物質に対する、すべての嗅覚受容体の応答を取得することは事実上不可能であり、使用可能な香気応答データは限定されている。そこで香気物質の分子構造パラメータに基づく嗅覚神経細胞の応答予測を行うことを考え、自己組織化写像(SOM)による嗅覚受容体の応答強度予測と変数選択法に関して検討をしたので本章にて述べる。

## 5.1 SOMによる非線形の回帰予測

SOMによる非線形の回帰予測の実装について説明する。SOMの学習は本来教師無学習として知られている。本研究では教師学習としてSOMを利用したので大まかな説明を本節で行う。‘観測値’が得られている入力ベクトルを教師信号として用い、‘観測値’+入力ベクトルからなるデータマトリックスPを形成する、そののち自己組織化写像(SOM)を計算する。そののち‘観測値’を求めようとする未知入力ベクトルに関して、求められた自己組織化写像の入力ベクトルに相当する部分と最も近くなるような参照ベクトルを抽出し、‘観測値’に相当するベクトル要素を求めようとしている未知の‘観測値’とする。この手法はすでに当研究室広田氏の2005年修士論文[090]の中で検討されている。

### 5.1.1 パラメータ空間の非線形マッピングを利用した予測手法

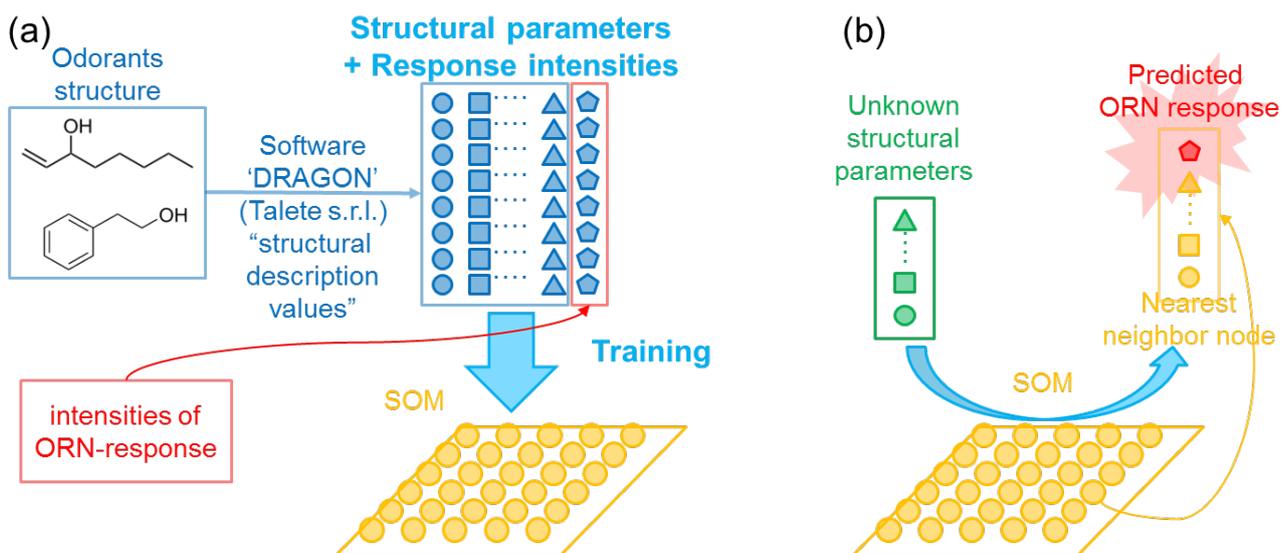


図 5-01.パラメータ空間の非線形マッピングを利用した予測手法

香気物質の構造を化学物質の構造のパラメータとして描述し、そのパラメータと注目する ORN から得られる応答強度を合わせて特徴ベクトルとする。このようにして得られたベクトル群(データテーブル)をもとに図 5-01(a)のように SOM のネットワークを学習させる。すると SOM 中の各ノードベクトルは応答強度に相当するパラメータを持つ構造となる。注目する ORN の応答強度を推測するためには、応答強度未知の香気物質を構造パラメータ化し、図 5-01(b)のようにその最近傍参照ベクトルを SOM から選択し、応答強度に対応するパラメータを出力とする。

## 5.2 6 種類の ORN (Or22a, Or35a, Or67a, Or47b, Or65a, Or88a)の 108 種類の香気物質に対する応答強度

**表 5-01. SOM を用いた嗅覚受容体の応答予測で対象とした 6 種類の ORN**

broadly tuned ORNs' response	Or22a
	Or35a
	Or67a
narrowly tuned ORNs' response	Or47b
	Or65a
	Or88a

まず自己組織化写像(SOM)による嗅覚受容体の応答強度予測が可能であるのかについて検証してみた。Hallem らの報告した 110 種類の香気物質に対する応答が調査された 24 種類の ORN のうち、6 種類の ORN 応答に対して、SOM を用いた嗅覚受容体の応答予測を行った。表 5-01 に示す比較的狭い分子種に対して応答する ORN3 種類(Or22a, Or35a, Or67a)と、比較的広い分子種に対して応答する ORN3 種類(Or47b, Or65a, Or88a)の合わせて 6 種類の ORN の応答を使用した。Talete DRAGON を用いた分子構造パラメータを得る過程で、methanol と ammonium hydroxide は得られる有効パラメータが著しく少なかったため検討対象から除外し、表 5-02 に示す 108 種類の香気物質に対して、結果は交差検定に基づき評価した。

### 5.2.1 自己組織化写像(SOM)による嗅覚受容体の応答強度予測

分子構造パラメータを削除したり圧縮したりせずに、変動していないパラメータや NaN の形でエラーが出ていないすべての分子パラメータを使用した。全てのパラメータ毎にオートスケーリングを行い(平均 0、標準偏差 1 に規格化)、(108 種類の香気物質) × (代表的な 6 種類の ORN の応答強度)に対して、“leave-one-out”法を用いて、SOM から推測される予測値の確からしさを検証した。その結果は図 5-02 にヒストグラムの形で示す。ヒストグラムに合わせて、複数通り試行した異なる SOM のサイズ、および交差検定(leave-one-out)における得られた値と元の値の間の標準偏差を算出した。

表 5-02. SOM を用いた嗅覚受容体の応答予測で対象とした 108 種類の香気物質

1	putrescine	37	nerol	73	1-octanol
2	cadaverine	38	linalool	74	2-pentanol
3	g-butyrolactone	39	b-citronellol	75	3-methylbutanol
4	g-hexalactone	40	linalooloxide	76	3-methyl-2-buten-1-ol
5	g-octalactone	41	acetaldehyde	77	1-penten-3-ol
6	g-decalactone	42	propanal	78	1-octen-3-ol
7	d-decalactone	43	butanal	79	E2-hexenol
8	methanoicacid	44	pentanal	80	Z2-hexenol
9	aceticacid	45	hexanal	81	E3-hexenol
10	propionicacid	46	E2-hexenal	82	Z3-hexenol
11	butyricacid	47	furfural	83	glycerol
12	pentanoicacid	48	2-propenal	84	2,3-butanediol
13	hexanoicacid	49	acetone	85	methylacetate
14	heptanoicacid	50	2-butanone	86	ethylacetate
15	octanoicacid	51	2-pentanone	87	propylacetate
16	nonanoicacid	52	2-heptanone	88	butylacetate
17	linoleicacid	53	6-methyl-5-hepten-2-one	89	pentylacetate
18	isobutyricacid	54	2,3-butanedione	90	hexylacetate
19	isopentanoicacid	55	phenethylalcohol	91	isobutylacetate
20	pyruvicacid	56	benzylalcohol	92	isopentylacetate
21	2-ethylhexanoicacid	57	methylsalicylate	93	E2-hexenylacetate
22	lacticacid	58	methylbenzoate	94	methylbutyrate
23	3-methylthio-1-propanol	59	ethylbenzoate	95	ethylbutyrate
24	dimethylsulfide	60	phenethylacetate	96	hexylbutyrate
25	terpinolene	61	benzaldehyde	97	ethyl3-hydroxybutyrate
26	a-pinene	62	phenylacetaldehyde	98	ethylpropionate
27	b-pinene	63	acetophenone	99	ethylmethanoate
28	(1S)-(+)-3-carene	64	ethylcinnamate	100	methylhexanoate
29	limonene	65	2-methylphenol	101	ethylhexanoate
30	a-humulene	66	4-ethylguaiaicol	102	hexylhexanoate
31	b-myrcene	67	eugenol	103	methyloctanoate
32	(-)-trans-caryophyllene	68	ethanol	104	ethyloctanoate
33	p-cymene	69	1-propanol	105	ethyldecanoate
34	geranylacetate	70	1-butanol	106	ethyltrans-2-butenate
35	a-terpineol	71	1-pentanol	107	ethylactate
36	geraniol	72	1-hexanol	108	diethylsuccinate

### 5.3 SOM サイズ

6 種類の ORN (Or22a, Or35a, Or67a, Or47b, Or65a, Or88a) の 108 種類の香気物質に対する応答強度の予測精度を交差検定によって SOM サイズを比較した。

以降の SOM 学習条件は SOM サイズ以外、4.5.1 節で使用した条件を用いた。学習回数は 100 回、初期近傍サイズは 3、SOM のレイヤーのユニットの配列のトポロジーは 'hextop' (平面 hexagonal 状 = ハニカム状 (六辺形) にユニットを並べる)、ニューロンの距離関数は 'linkdist' (=link distance) とし、そのほかのパラメータは neural network toolbox パッケージで提供される SOM コマンドのデフォルト値を用いた。

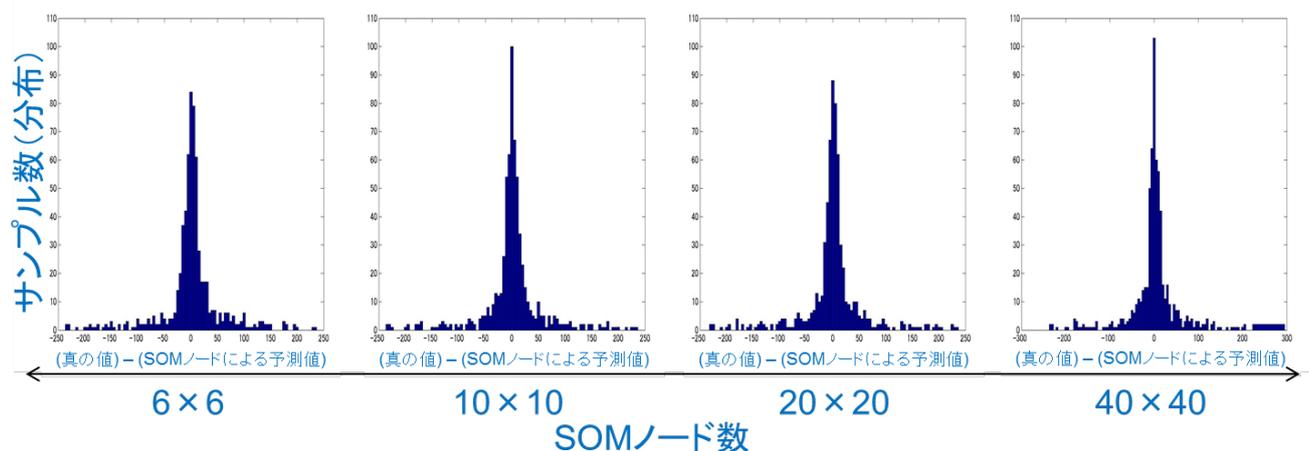


図 5-02. 応答予測(108 種類の香気物質) × (代表的な 6 種類の ORN の応答強度)における交差検定結果

表 5-02. 応答予測(108 種類の香気物質) × (代表的な 6 種類の ORN の応答強度)における交差検定結果

opSizeX	opSizeY	絶対値の平均	標準偏差
6	6	31.7303	55.1908
10	10	31.5855	56.7733
20	20	30.7521	54.7470
40	40	31.8146	56.4481

図 5-02 に示すように SOM のサイズとしては 6×6 から 10×10、20×20、40×40 を検証した。交差検定における得られた値と元の値の誤差絶対値平均、標準偏差を確認した結果を表 5-02 に示す。今回の検証において著しく良い結果を示した SOM サイズは結論できなかった。機械学習の手法中のパラメータ選択により傾向が変わる可能性も高いことから、ノードのサイズへの依存は検証できる段階ではないと考えられる。また図 5-02 のヒストグラムを確認するとわかるように、得られた値の中心値は真の値であったが、近い値が得られなかったものもあった。このような誤差の大きなサンプルの数を減らすことを主眼に、予測精度を改善してゆく必要があることが分かった。

#### 5.4 変数(分子情報)選択法

leave-one-out 法をもちいて、構造パラメータを削除したり圧縮したりせずに、すべての得られうるパラメータを使用して、ORN 応答強度推測ができることを確認し、自己組織化写像(SOM)を用いて、香気分子の化学構造からの嗅覚受容体の応答強度を予測できた。誤差の大きなサンプル例がそれなりにあることから、予測精度の改善が今後求められることが明らかになった。また Talete DRAGON を用いて得られる、それぞれの化合物に対する約 5000 種類の分子構造パラメータは多重共線性を持ち、分子パラメータの構造は冗長になっていると考えられる。これはマッピングを用いた予測において、精度の向上と安定化のために機械学習の手法中のパラメータ選択にはまだ改善の必要ある。

推測値をより精度の高いものとする事が必要であり、その為には得られうるすべての構造パラメータを使用するのではなく、選択された構造パラメータを使用することが有用だと考えられた。

### 5.4.1 条件数に基づく選定

構造パラメータの選択においては、これまでに当研究室において検討された条件数による選択が有用であると考えている。抽出されたパラメータテーブルを  $G$  と表記すると、次の式で示すような条件数が定義できる(図 5-03)。ここで  $Q$  は対角成分にのみ値を持つ  $G$  の規格化行列である( $G$  の各行の総和が一定になるように規格化する)。これが最小となるようなパラメータのみを使用することで、精度のよい効率的な応答強度の推測が可能になると考えられる(図 5-04)。

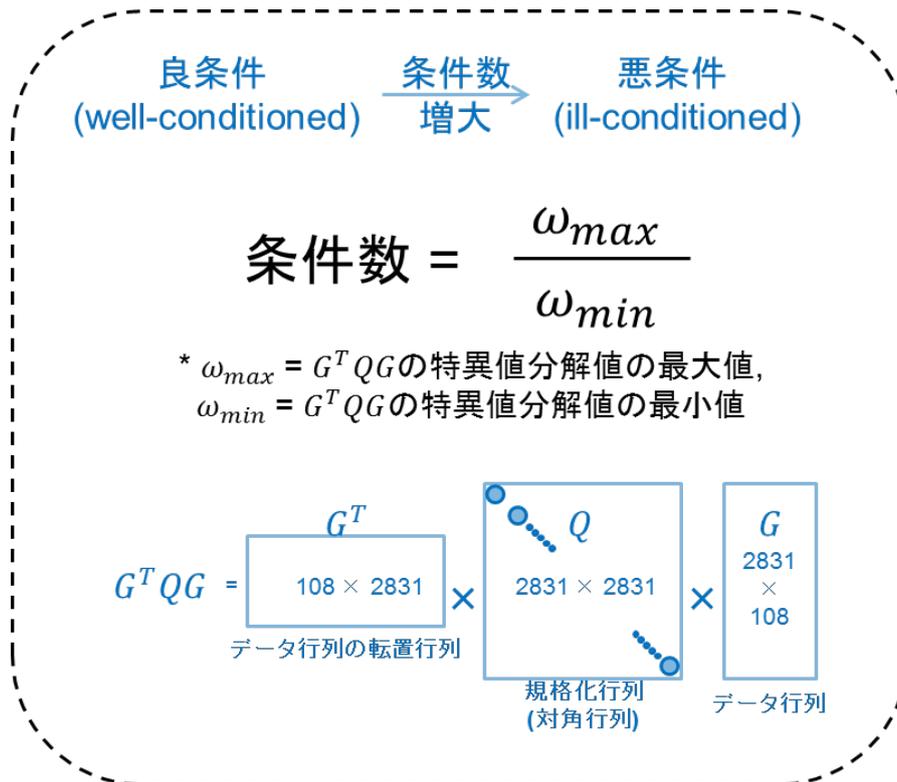


図 5-03. 条件数の定義

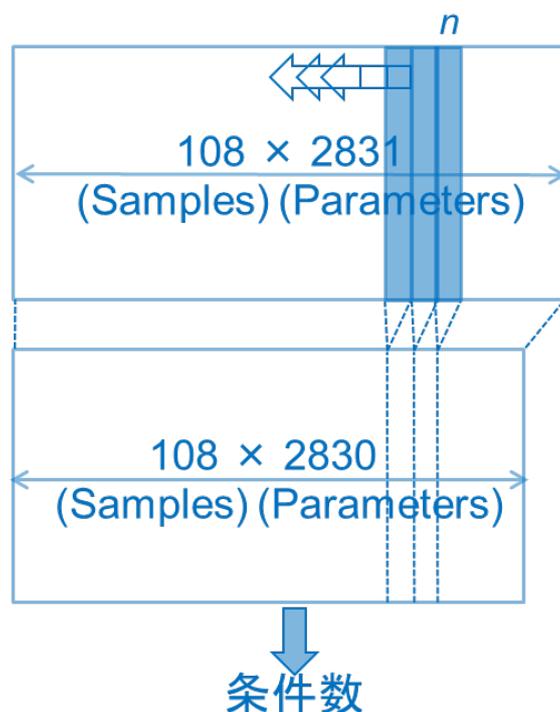


図 5-04. 条件数への影響度の決定方法

まずパラメータテーブルに関して、変動していないパラメータや NaN が出力に含まれていたパラメータを排除して、パラメータ数 2831 個のデータに成型した。そののちに条件数への影響度に分子構造パラメータを並べ替えた。図 5-04 のように、条件数への影響度を検証した。

1. 2831 パラメータの中から一つのパラメータを選択
2. そのパラメータを取り除いた場合について、条件数を算出。
3. 全パラメータ組合せに対してこのような評価を繰り返し、全パラメータの条件数への寄与を序列化

条件数への影響度に並べ替えた分子構造パラメータの最適パラメータ数の見積もりを得たいと考え、この序列の中の順に精度寄与の低かったパラメータから順番に取り除きながら都度条件数を計算させ、プロットしたのが、図 5-05 である。この図からはパラメーターがかなり減っても条件数が上昇しないことがわかる。ある一定のところから条件数の上昇が始まり、分子構造パラメータの情報空間において効率的なパラメータ数量がある事が示唆された。変数(分子情報)を選択することで、精度の向上が見込めるのではないかと考察する。

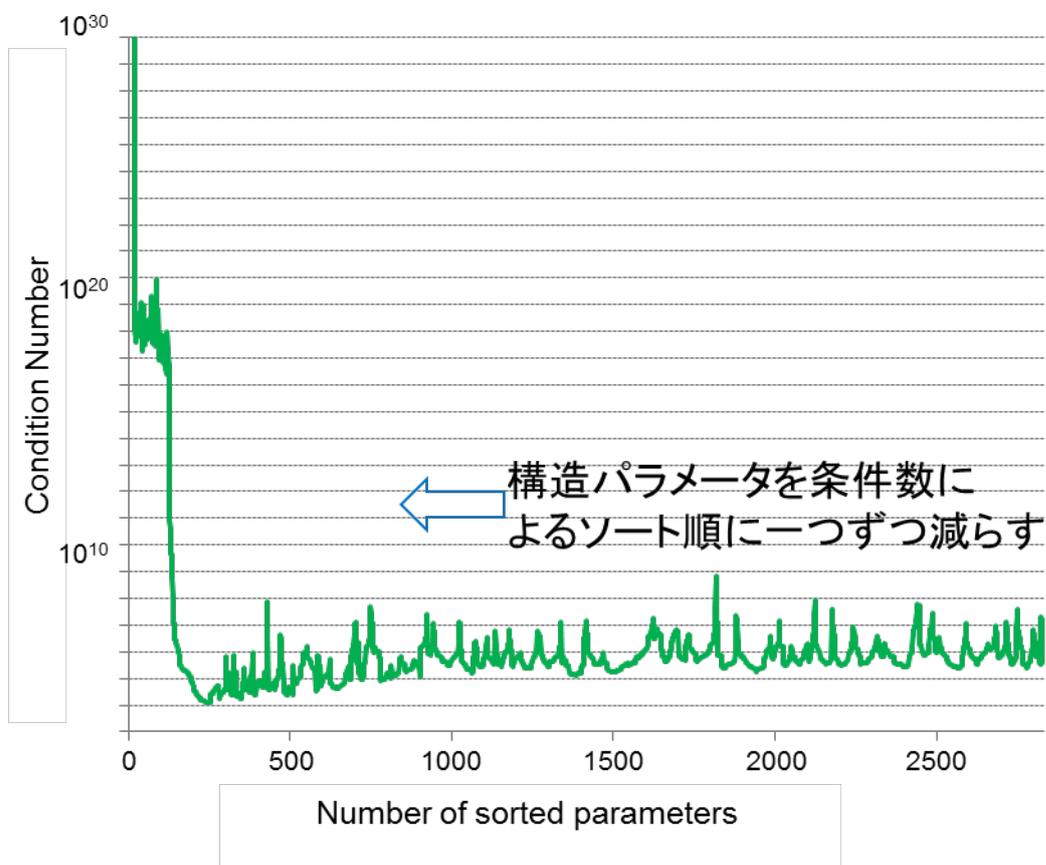


図 5-05.条件数への影響度に並べ替えた分子構造パラメータの良組合せ数の見積もり

#### 5.4.2 使用するパラメータ個数と条件数の関係

パラメータ最適組み合わせ探索から、SOM を用いた ORN 応答強度の予測において効率的なパラメータ数量がある事が示唆された。大まかに変数(分子情報)を選択することで、条件数への影響度に並べ替えた分子構造パラメータの、精度の向上への影響度を交差検定で検証した。

#### 5.4.3 ‘AIC’を用いたパラメータ選択(使用するパラメータ数)の検討

これまでに構造パラメータの数を減らして行くと SOM による ORN 応答強度の標準偏差(以降、表中などで SSTD として示す)など推測精度が悪化することがわかっている。ただし推測の結果が良い反面、多すぎる構造パラメータの利用は、多重共線性の問題を持つために、パラメータの多さの割に良好な予測精度を持たないと推測される。そのため、SOM による予測精度の性能評価の手法として AIC(赤池情報量規準、Akaike's Information Criterion)を利用することにした。

AIC は以下のような式を用いて算出する。

$$AIC = n \log(2 \times \pi + 1) + n \log \frac{\sum_{i=1}^n ((\text{true value})_i - (\text{predict})_i)^2}{n} + 2 \times DF + 4 \quad (\text{式 5.1})$$

赤池情報量規準(AIC)は、統計モデルの良さを評価するための指標として知られる n はサンプル数を表す、DF は自由パラメータの数を表し、本検討では使用した分子構造パラメータの数をを用いた。交差検定を AIC により評価し、最適なパラメータを選定した。

6 種類の ORN (Or22a, Or35a, Or67a, Or47b, Or65a, Or88a) の 108 種類の香気物質に対する応答強度の予測精度を交差検定によって確認し、分子構造パラメータを条件数への影響度を指標に大まかに変数(分子情報)を選択することの、精度の向上への影響度を交差検定を用い、図 5-06(a)に示すように検証した。

- 条件数により見積もられた影響度に基づき、上位から 30 種類～500 種類のパラメータを選択
- “leave-one-out”法をもちいて(108 種類の香気物質) × (代表的な 6 種類の ORN の応答強度)、SOM から推測される予測値の確からしさを検証した。
- SOM サイズをいくつか変更しながら、比較検討した。

**表 5-03.得られた予測値と実測値をもとにした AIC**

SOM size \ number of parameter		30	50	100	200	300	500	2831
6 × 6	AIC	6657	6618	6726	6887	7103	7492	12151
	標準偏差	60.0	56.5	56.8	55.1	55.8	55.4	55.2
10 × 10	AIC	6704	6668	6743	6918	7074	7525	12187
	標準偏差	62.2	58.7	57.6	56.5	54.6	56.8	56.8
20 × 20	AIC	6693	6665	6776	6948	7131	7484	12141
	標準偏差	61.7	58.6	59.1	57.8	57.1	55.0	54.8
40 × 40	AIC	6739	6750	6769	6996	7250	7566	12180
	標準偏差	63.8	62.5	58.7	60.0	62.5	58.6	56.5

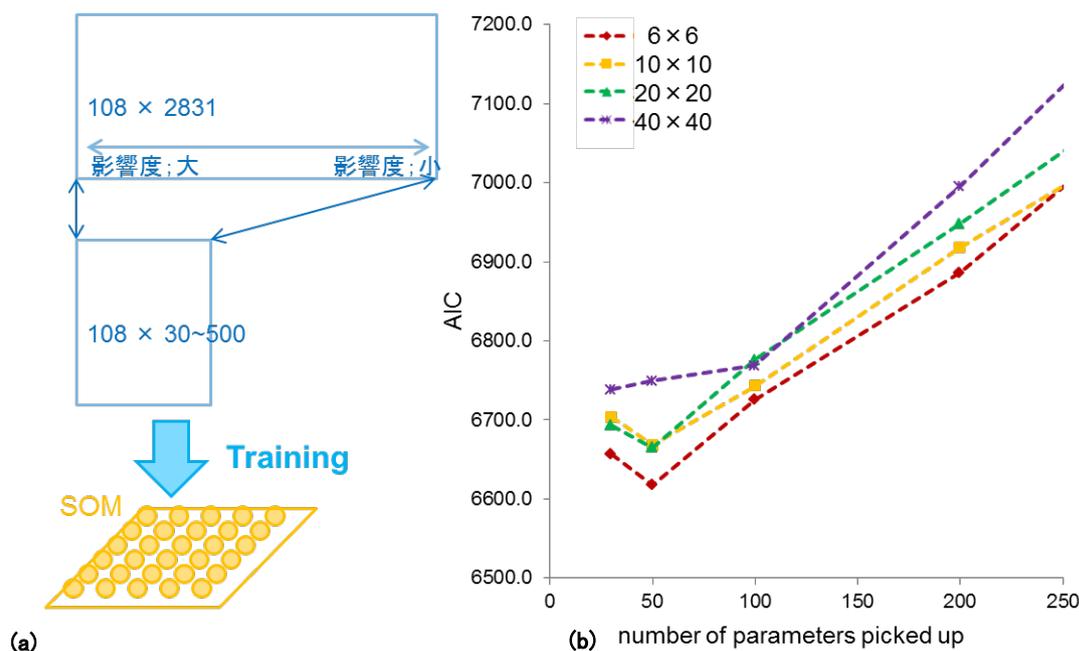


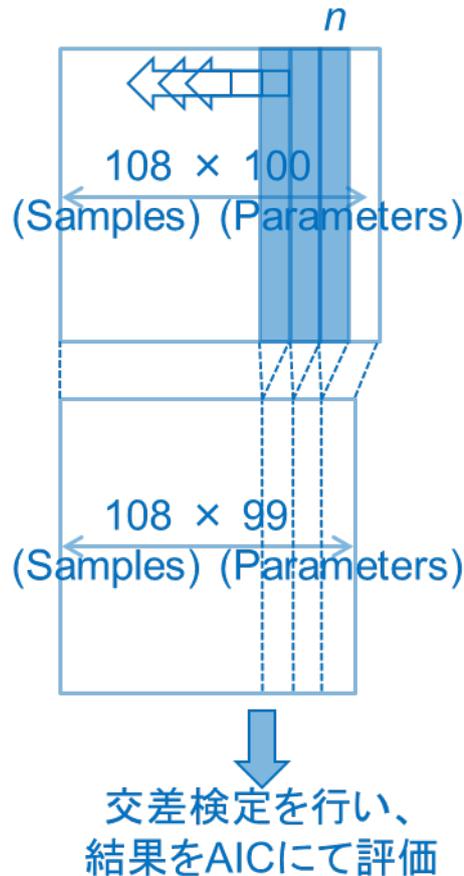
図 5-06. AIC による嗅覚受容体の応答強度予測の性能評価 ; (a) SOM による応答強度予測に利用したパラメータの選択方法の模式図、(b)表 5-03 に示す AIC の使用したパラメータ数に対するプロット

分子構造パラメータを条件数への影響度順に並び替え、影響度の大きい方から 30~500 個抜き出した。それぞれ選択したパラメータを用いて図 5-06(a)に模式的に示すように、SOM による応答強度予測を行った場合の、leave-one-out 法による交差検定の結果を表 5-03 に示す。すべての香気種類に対する leave-one-out 法を試行し、SOM-ORN 応答強度予測から、ORN の予測値と実際の値を対比させて、AIC および標準偏差で評価した。図 5-06(b)に示すのは、予測値と実測値に基づいて得られた表 5-03 にも示した AIC を使用したパラメータ数に対してプロットしたものである。この図から使用したパラメータ数 50 個のとき、ほとんどの SOM サイズで AIC が最も低くなることが分かった。この検討から SOM を用いた ORN 応答強度の予測において効率的なパラメータ数がある事が示唆された。

またこの検討でも、SOM のサイズとしては 6×6 から 40×40 までを検証したが、最適な SOM サイズは結論できず、ノードのサイズへの依存は検証できる段階ではないと考えられる。

#### 5.4.4 交差検定の AIC による評価に基づく最適なパラメータの選定

これまでに条件数をもとにしたパラメータの絞り込みをした上で、交差検定による SOM 形成による構造パラメータの絞り込みをしようとしてきたが、条件数計算そのものとは別の指標を入れなくては最適パラメータ組み合わせを導出することはできない。そのために、図 5-07 に示すような条件数計算である程度絞り込みをしたパラメータ群に対して AIC を用いて直交性の高い組み合わせを選択する手法を検証する。



**図 5-07. SOM による応答強度予測に最適パラメータ組み合わせを導出する方法**

上記の検討からは実際にパラメータ数が適正かどうか、選択したパラメータ組み合わせが適正かどうかは保証されない。そこで、100 種類のパラメータを選択し、予測精度に影響度が少ない方からパラメータを減らしながら最適なパラメータを選定することにした。(SOM サイズは  $6 \times 6$ )

1. あらかじめ条件数検討から 100 パラメータまで削減、その中から一つのパラメータを選択
2. そのパラメータを取り除いた場合について、交差検定を用いて新しいパラメータ組み合わせにおける予測性能を評価する。
3. 全パラメータ組合せに対してこのような評価を繰り返し、全パラメータの予測精度への寄与を AIC にて序列化
4. この序列の中の最も精度寄与の低かったパラメータを取り除き、選択パラメータ群に加える。そして1~3を同様に繰り返す。

選ばれたパラメータの中から一つのパラメータを選択し、そのパラメータを取り除いた場合について、交差検定を用いて新しいパラメータ組み合わせにおける予測性能を評価する。全パラメータに対してこのような評価を繰り返し、全パラメータの予測精度への寄与を序列化した。この序列の中の最も精度寄与の低かったパラメータを取り除き、選択されたパラメータ群に加えた。

#### 5.4.5 提案する最適各パラメータ組合せ探索法

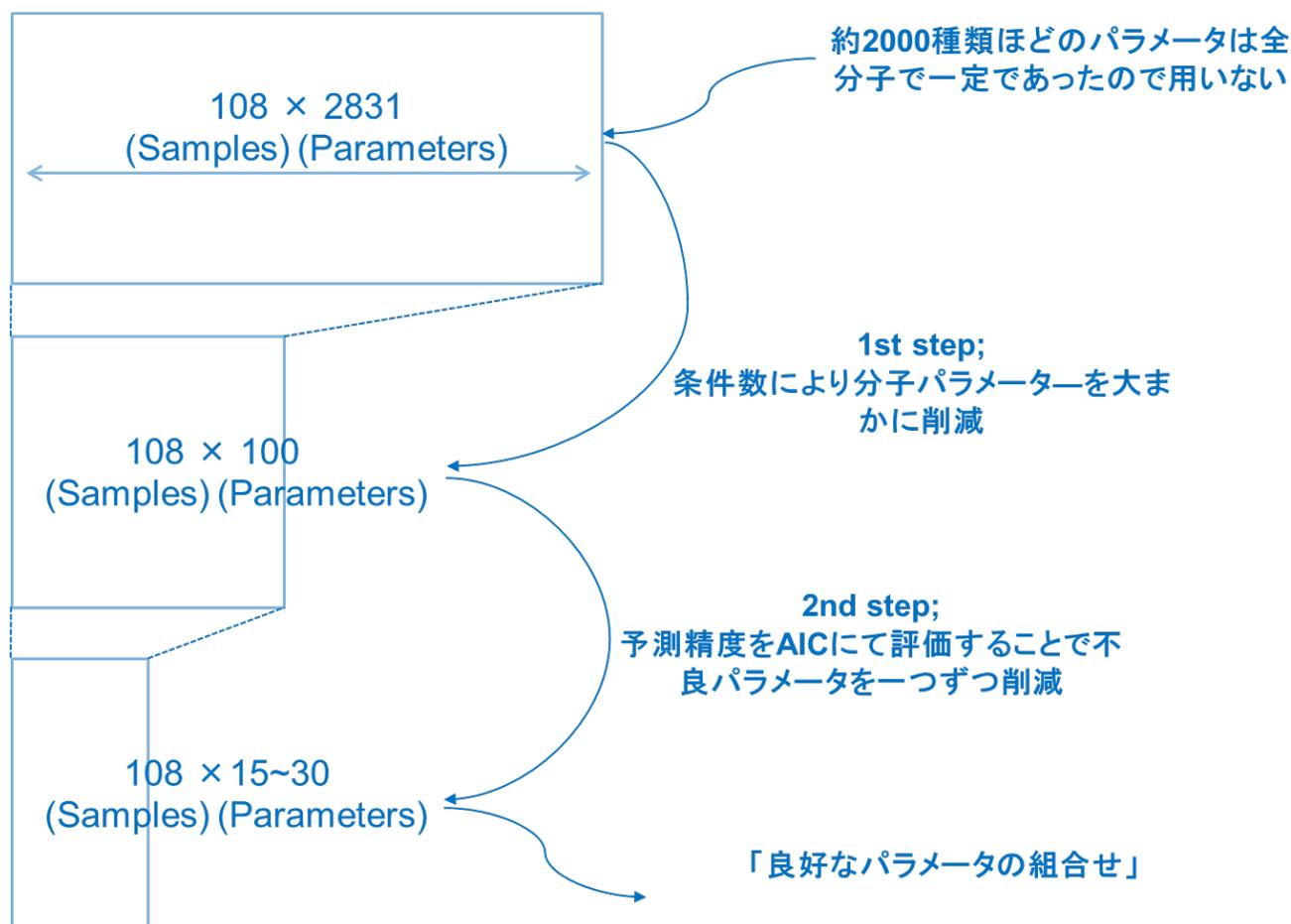


図 5-08. 提案する最適変数(分子構造パラメータ)選択手法

前節にて説明したように、変数(分子情報)を選択することで精度の向上が見込めるのではないかと考え、2段階のパラメータ削減手法を提案した。まずパラメータの組み合わせにおける条件数の比較から大まかにパラメータを削減し、交差検定の AIC に基づき、パラメータを掃き出し法により一つずつ減らすことで最適なパラメータ組合せを得ようとした。条件数計算と AIC による評価をもとに、パラメータを漸減させながら最適各パラメータ組合せにより応答予測を行う手法を提案した。

最適パラメータ組合せのために用いた交差検定では、計算速度向上のために leave-one-out 法ではなく、ランダムな交差検定用の 25% のデータの選択を行い、残り 75% を学習に用いた。この検定を交差検定用データを変えながら 8 回行い、AIC の算出や真値と予測値の比較は 8 回分の交差検定を合わせたものに基づいた。以降、別途説明する場合を除き、最適パラメータ組合せのために用いた交差検定では、以上の手法を用いる。

## 5.5 最適各パラメータ組合せ探索法を用いた自己組織化写像(SOM)による嗅覚受容体の応答強度予測

### 5.5.1 電気生理学的測定の結果に対する自己組織化写像(SOM)による嗅覚受容体の応答強度予測

Hallem らの論文及び DoOR database にて報告されている OR65a, OR13a に対して、最適各パラメータ組合せ探索法を用いた自己組織化写像(SOM)による嗅覚受容体の応答強度予測を適用した。

自己組織化写像(SOM)による嗅覚受容体の応答強度予測の対象として ORN の電気生理学的測定の結果 (OR65a @Hallem)を使用した。最適なパラメータの選定による予測手法を適用した結果を図 5-09、図 5-10 に示す。OR65a は香気物質に対する選択性が比較的狭い香気物質に応答する受容体である。パラメータ数がおおよそ 10 個の時に、交差検定の AIC 評価が最小となった。パラメータを漸減する過程での AIC 値は図 5-09 のようにプロットされる。その条件における交差検定結果の実測値からの誤差をヒストグラムの形であらわすと図 5-10(a)のようになる。図 5-10(b)に示すのは交差検定の結果を横軸に実際の値、縦軸に予測された応答強度でプロットしたものである。これらの図にて確認したところ、予測精度は比較的高かった。

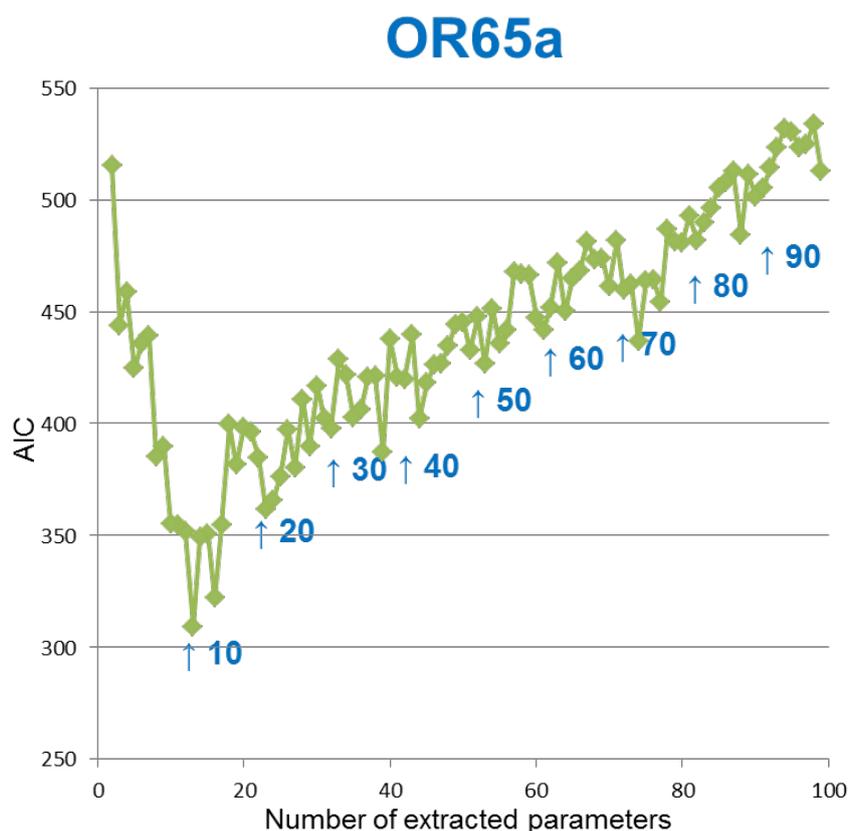


図 5-09. 2 段階の変数選択を使用した SOM による ORN 応答予測(対象 OR65a @Hallem) における交差検定の最良 AIC プロット

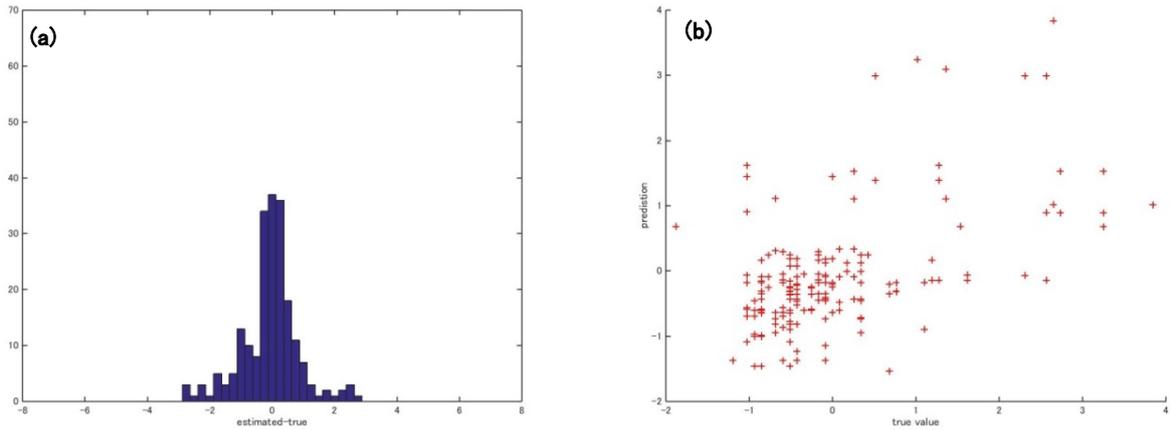


図 5-10. 提案した 2 段階の変数選択を使用した SOM による ORN 応答予測(対象 OR65a @Hallem) ;最適なパラメータの選定により 10 個のパラメータを選択したときの交差検定の結果 ;(a)はその実測値を横軸に予測値を縦軸にとったプロット、(b) は OR65a @Hallem に対する誤差のヒストグラム、(b)はその実測値を横軸に予測値を縦軸にとったプロット

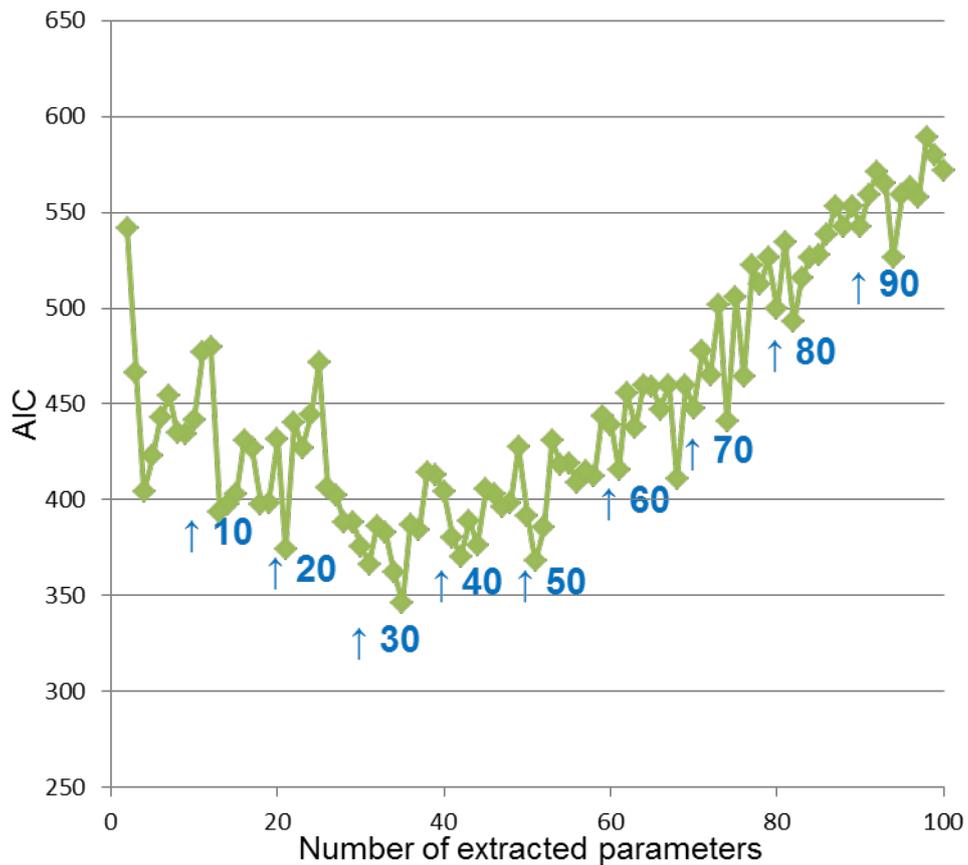
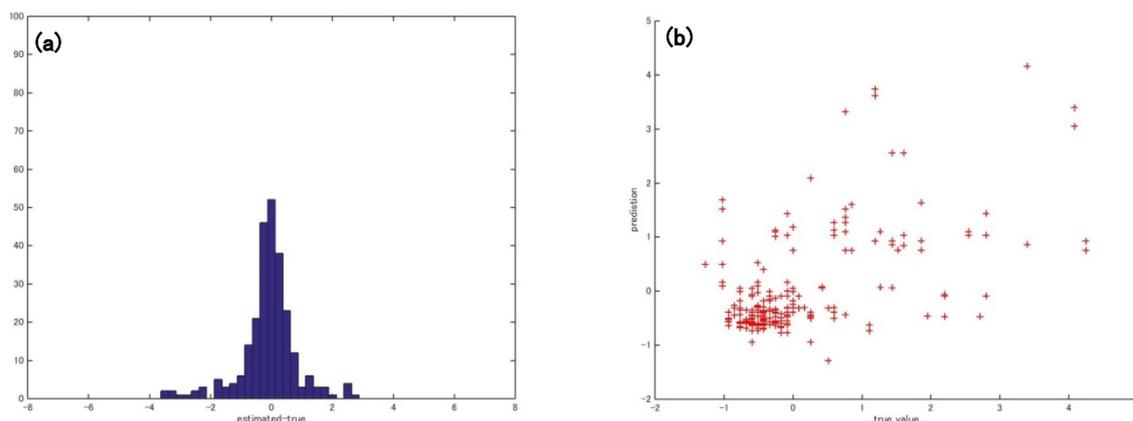


図 5-11. 提案した 2 段階の変数選択を使用した SOM による ORN 応答予測(対象 OR13a @DoOR)における交差検定の最良 AIC プロット



**図 5-12. 提案した 2 段階の変数選択を使用した SOM による ORN 応答予測(対象 OR13a @DoOR) ;最適なパラメータの選定により 30 個のパラメータを選択したときの交差検定の結果 ;(a)はその実測値を横軸に予測値を縦軸にとったプロット、(b)は OR13a に対する誤差のヒストグラム、(b)はその実測値を横軸に予測値を縦軸にとったプロット**

次に自己組織化写像(SOM)による嗅覚受容体の応答強度予測の対象として ORN の電気生理学的測定の結果 (OR13a @DoOR)を使用した。同様に最適なパラメータの選定による予測手法を適用した結果を図 5-11、図 5-12 に示す。OR13a は香気物質に対する選択性が比較的多くの香気物質に応答する受容体である。Hallem らの論文[060]では OR13a の応答は報告されていないが、現在彼らの研究グループのウェブサイト[091]にてデータを閲覧できる。パラメータ数がおよそ 30 個の時に、交差検定の AIC 評価が最小となった。パラメータを漸減する過程での AIC 値は図 5-11 のようにプロットされる。その条件における交差検定結果の実測値からの誤差をヒストグラムの形であらわすと図 5-12(a)のようになる。図 5-12(b)に示すのは交差検定の結果を横軸に実際の値、縦軸に予測された応答強度でプロットしたものである。これら図にて確認したところ、予測精度はそれなりに高く、OR65a @Hallem をターゲットとして使用した時よりも良好な結果を得た。

OR13a は香気物質に対する選択性が比較的幅広い香気物質に反応する(ブロードな選択性)受容体である。ブロードな選択性を持つ受容体の応答予測に対して有効であることが示唆された。

### 5.5.2 香気物質の空間に基づく、能動学習のための検討

SOM はニューラルネットの一種で、多次元データを圧縮して低次元のマッピングし、情報構造の可視化のツールとなる。4.5.2 節にて説明した手法と同様に、印藤 元一「合成香料」[088]をもとに、得られた Tallete DRAGON によるパラメータを各パラメータ毎にオートスケーリング(標準偏差 1 平均 0 になるように規格化)し、市販の約 1500 種類の香料化合物の分子構造をもとに化合物群のマッピングを得た。このように、物質の写像を調査することにより、注目している物質と似ている物質・異なる物質を探すことが可能である。

香料空間中に投射された物質の番号は黒で、Hallem の香気番号の色を変えてその分散を可視化した。図 4-09 中において、学習後の SOM のノード間距離を近い方から黄色→オレンジ→赤→黒で示した。分子構造を元に、分子構造パラメータ空間への写像を調査した。このように香気物質の分子構造を考慮し、より多くの応答情報

を取得することで、全体的な受容体の応答予測の精度を向上させられることが期待できる。

OR65a, OR13a のデータへ応答予測法を適用したところ、OR13a のデータについて比較的良好な予測が得られたことを前節にて述べた。この SOM を用いた ORN 応答予測における交差検定の結果をもとに、約 1500 種類の香気物質の番号をマッピングしたうえで、香気物質の空間に基づく能動的なデータ収集手法を検討した。Hallem らの実験に用いられた 108 種類の化合物に関して、良好な予測が得られる化合物とそうでない化合物を調査した。図 5-13、図 5-14 中において予測結果が良好だった香気物質を青色の番号で、予測結果が不良だった香気物質を赤色の番号で示す。予測精度の低いサンプルの周辺データを追加して、学習することで、全体的な予測精度の向上が期待される。このように学習結果に応じて学習サンプルを適応的に変えていく能動学習法の開発も今後の課題である。

予測結果が良好だった香  
 気物質を青色の番号  
 予測結果が不良だった香  
 気物質を赤色の番号  
 で示す  
 (対象: OR65a @Hallem)

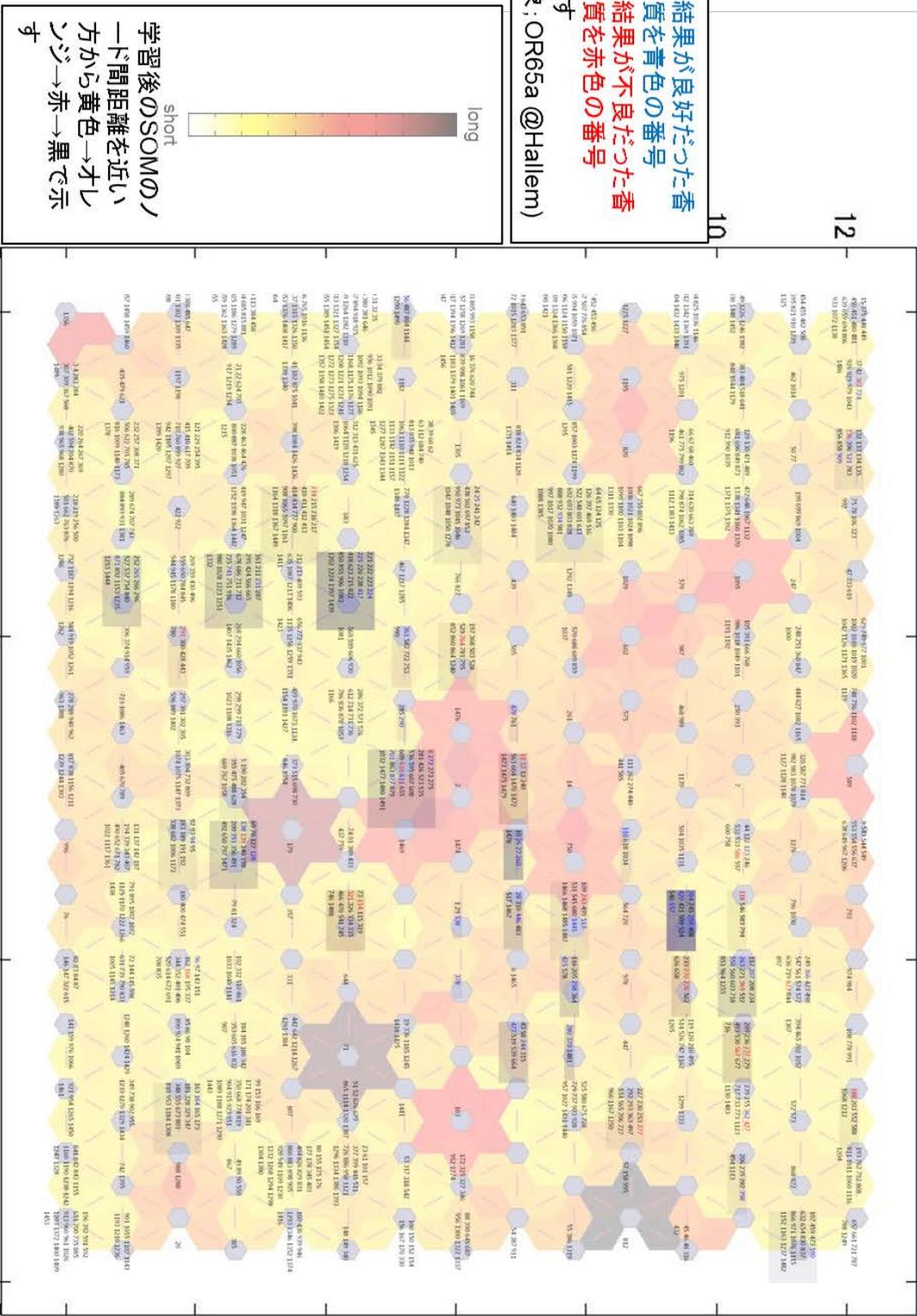


図 5-13. 予測(対象: OR65a)に用いた香気物質の市販の香料の化合物空間における解釈

予測結果が良好だった  
香気物質を青色の番号  
予測結果が不良だった  
香気物質を赤色の番号  
で示す  
(対象: OR13a@DOOR)

学習後のSOMの  
一ド間距離を近い  
方から黄色→赤→黒で示  
す

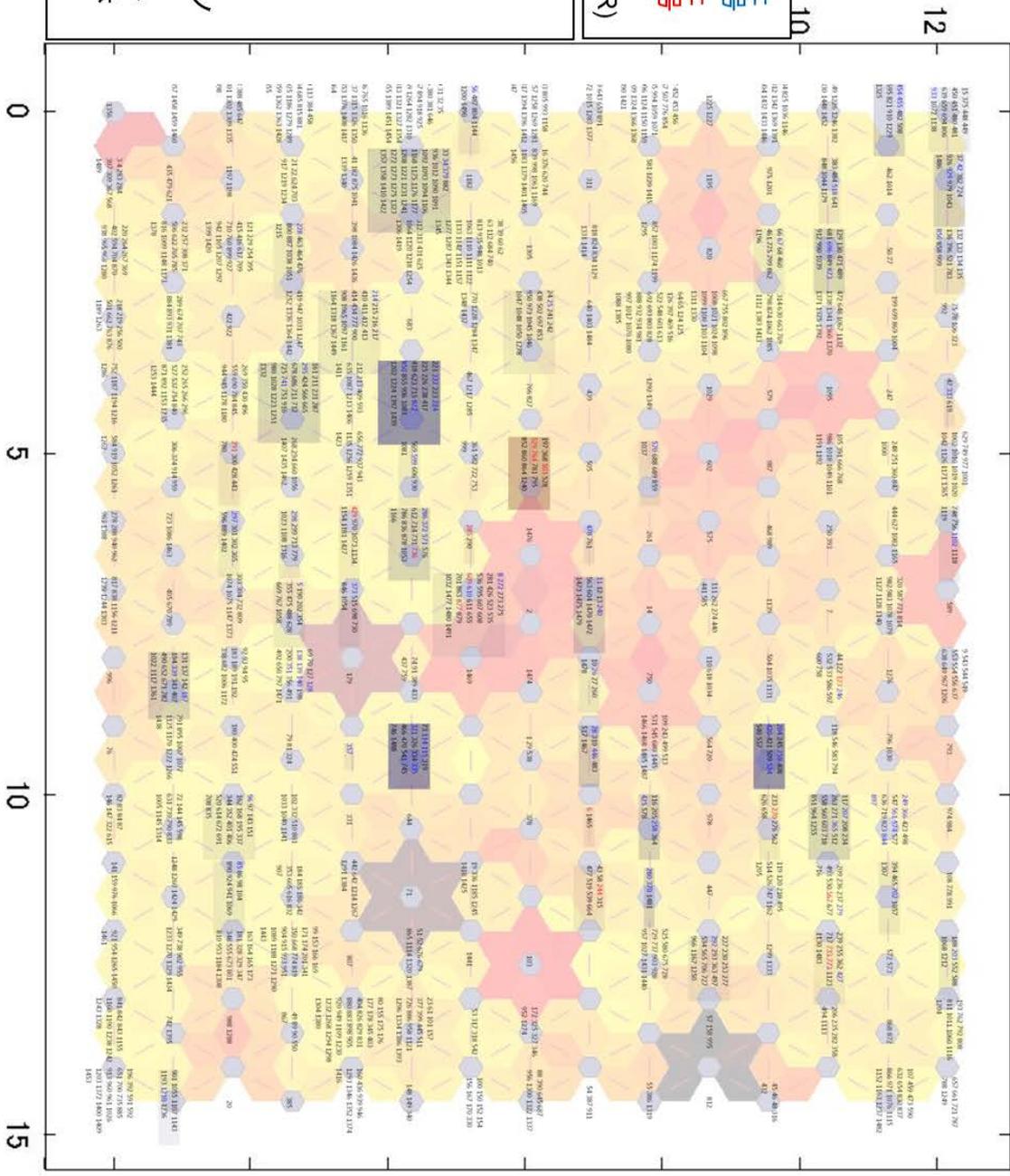


図 5-14. 予測(対象: OR13a)に用いた香気物質の市販の香料の化合物空間における解釈

## 5.6 主成分分析 (PCA)による香気分子の構造パラメータ圧縮とそれを用いた予測

### 5.6.1 主成分分析 (PCA)による香気分子の構造パラメータ圧縮

これまでに条件数をもとにしたパラメータの絞り込みをした上で、交差検定による SOM 形成による構造パラメータの絞り込みをしようとしてきたが、データを 2 章で述べた主成分分析 (PCA) によって解析することでも 108 次元まで構造パラメータが圧縮できる。本節での検討ではブロードな応答特性を持つ ORN の一例として OR13a ではなく OR67a のデータを Hallem らの論文[060]より引用して使用した。図 5-15 に示すのは主成分分析 (PCA) による解析の概念である。今回使用した香気物質は 108 種類であるので、主成分分析により第一主成分から第 108 主成分まで計算できる。元のデータを  $G$  は、圧縮されたパラメータ行列  $PCs$  と係数行列  $coef.$  の積の形で表現できる。この圧縮されたパラメータ  $PCs$  は合成パラメータであるものの、今回は主成分軸の抽象的な意味の検討は行わなかった。



図 5-15. 主成分分析 (PCA)による解析の概念

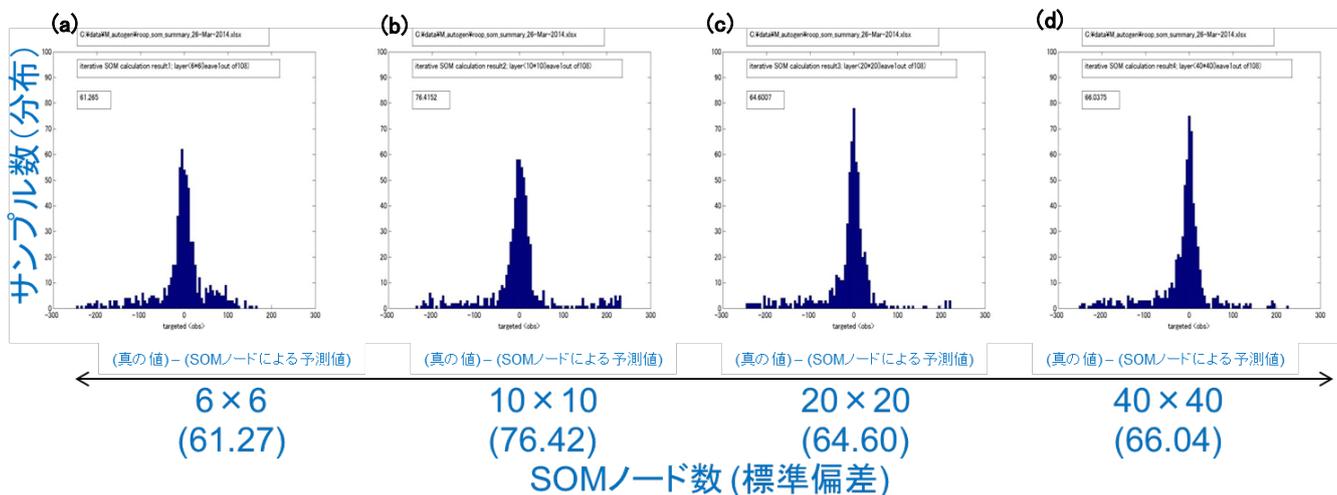
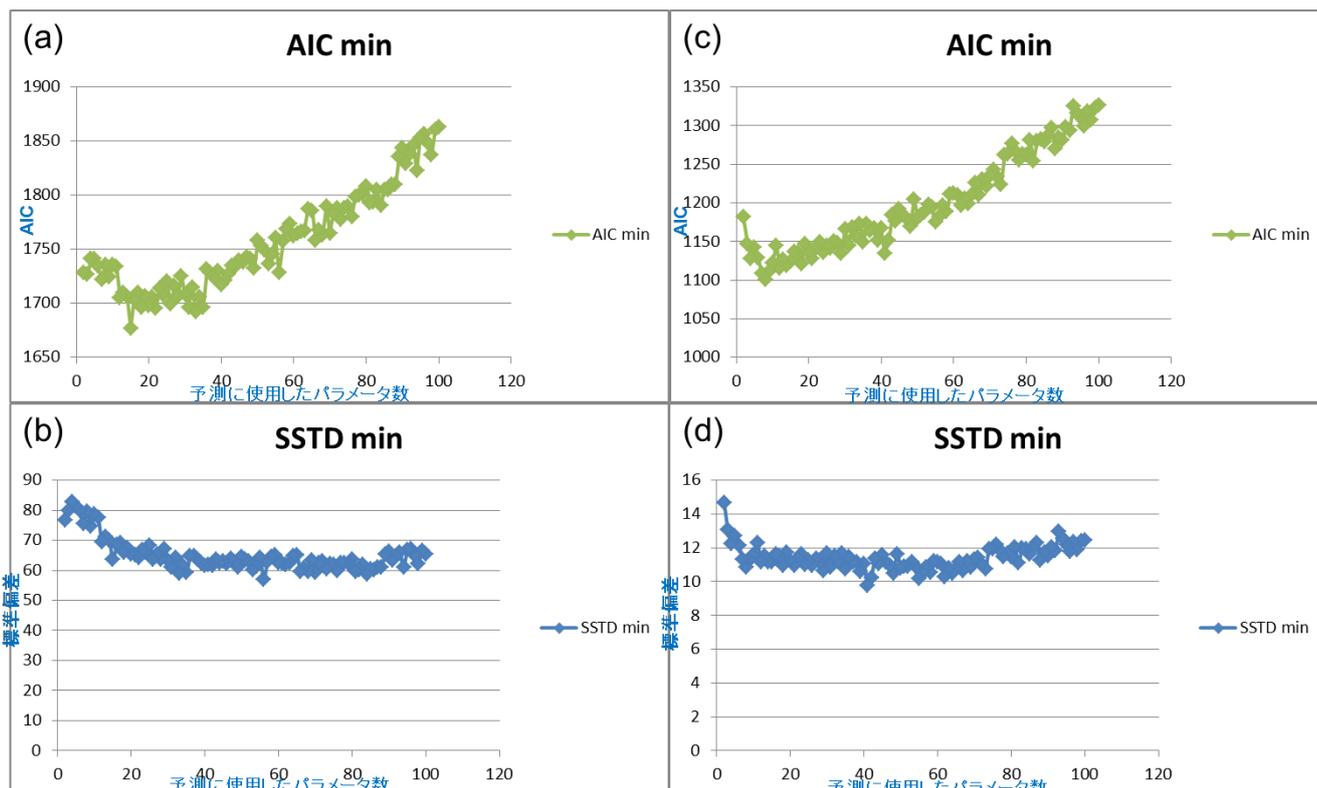


図 5-16. 主成分分析 (PCA)により圧縮された香気分子の構造パラメータを用いた応答予測における交差検定結果; (108 種類の香気物質) × (代表的な 6 種類の ORN の応答強度)

なお、この主成分データベースを用いて、“leave-one-out”法をもちいて、SOM から推測される予測値の確からしさの検証をした。表 5-02 に示す 108 種類の香気物質をもとに主成分分析 (PCA) により圧縮された香気分子の構造パラメータを用い、表 5-01 に示す代表的な 6 種類の ORN の応答強度の予測を行った。SOM のサイズとして

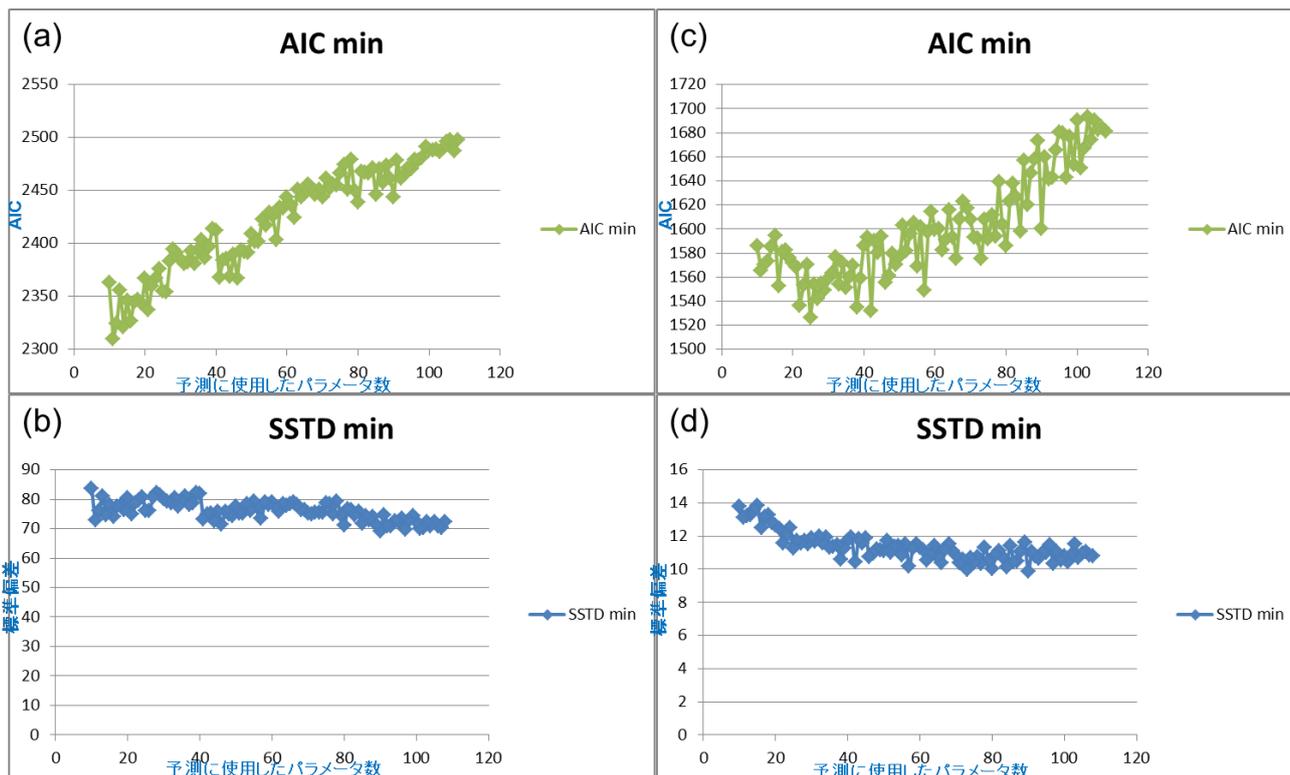
は 6×6 から 40×40 までを検証し、図 5-16 に交差検定における予測結果を実際の値との差のヒストグラムで示す。交差検定における得られた値と元の値の間の標準偏差を図 5-16 に ( ) 内に示した。この結果はすべてのパラメータをそのまま使用したときの検討において得られた標準偏差(sstd = 55.233, 56.817, 54.789, 56.492)と比べ、標準偏差は大きくなった。最適な SOM サイズはここでも結論できなかった。また、大幅に真の値からずれているサンプルが依然多い傾向も変わらない結果となった。

### 5.6.2 主成分分析(PCA)と変数選択を使用した SOM による ORN 応答予測



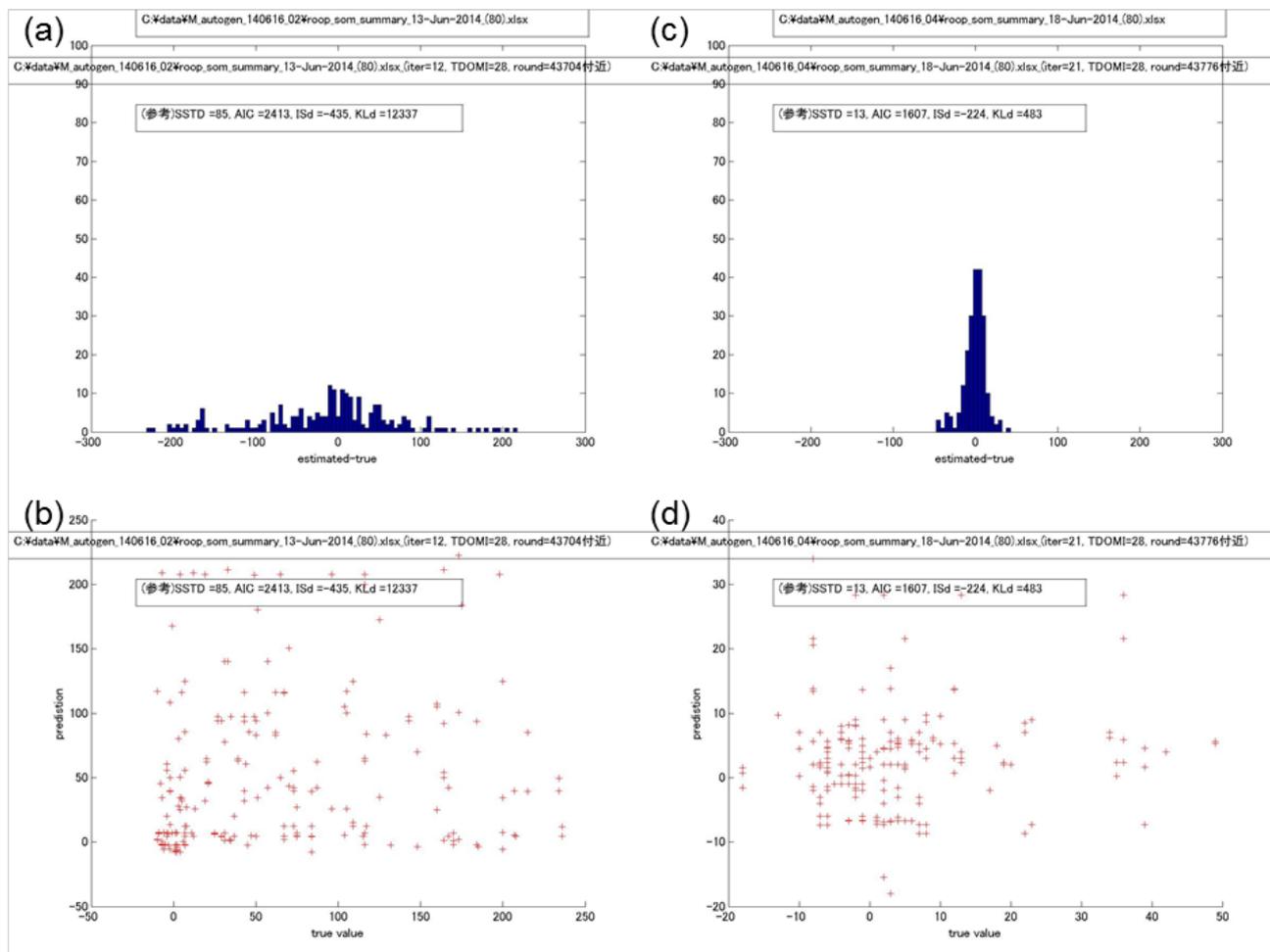
**図 5-17. 提案した 2 段階の変数選択を使用した SOM による ORN 応答予測(条件数に基づき事前に選択したパラメータにおいて変数選択を使用した, 対象 OR67a, OR65a) (a) は OR67a に対して手法を適用したときの AIC プロット、(b)はその AIC に対応する標準偏差、(c) は OR65a に対して手法を適用したときの AIC プロット、(d)はその AIC に対応する標準偏差**

主成分分析で香気分子の構造パラメータを圧縮できた。これによって得られたパラメータを元に SOM を用いた刺激応答予測を行うと、圧縮前よりも予測精度は若干悪化した。ただし、不要パラメータを順に除去しながら良好パラメータ組み合わせのみにしたとき、予測精度を向上させられるのか興味を持たれた。条件数による事前パラメータ絞込みの後に、5.4.4 章にて説明した不要パラメータを順に除去ステップを用いて、良好パラメータ絞りこみを行い、OR67a, OR65a @Hallem の応答に対する PCA で圧縮した合成分子パラメータの変数選択を使用した SOM による ORN 応答予測の精度を求めた。結果を図 5-17 に示す。



**図 5-18. PCA で圧縮した合成分子パラメータの変数選択を使用した SOM による ORN 応答予測(対象 OR67a, OR65a)、(a) は OR67a に対して手法を適用したときの AIC プロット、(b)はその AIC に対応する標準偏差、(c) は OR65a に対して手法を適用したときの AIC プロット、(d)はその AIC に対応する標準偏差**

図 5-18. に示すのは OR67a, OR65a の応答に対して、検討した PCA で圧縮した合成分子パラメータの変数選択を使用した SOM による ORN 応答予測の結果である。図 5-18(a) は OR67a に対して手法を適用したときの AIC プロット、図 5-18(b)はその AIC に対応する標準偏差、図 5-18(c) は OR65a に対して手法を適用したときの AIC プロット、図 5-18(d)はその AIC に対応する標準偏差を表す。



**図 5-19. PCA を用いて圧縮した分子構造パラメータを用い、2 段階の変数選択を使用した SOM による予測における、交差検定の結果(対象 OR67a, OR65a) ; (a) は OR67a に対して手法を適用したときの誤差のヒストグラム、(b)はその実測値を横軸に予測値を縦軸にとったプロット、(c) は OR65a に対して手法を適用したときの誤差のヒストグラム、(d)はその実測値を横軸に予測値を縦軸にとったプロット**

予測に使用する分子構造パラメータを減らしてゆくと、定のパラメータ数以下で AIC が上昇することがわかる。標準偏差もほぼ同じパラメータ数以下になったとき上昇を始めていると考察できる。それに対して図 5-18 では一定のパラメータ数以下での AIC の上昇、標準偏差の上昇は見られるものの、変曲点はわかりにくいものになっている。SOM を用いた ORN 応答強度の予測において効率的なパラメータ数量がある事が示唆された、と考察できる。

全体的な AIC・標準偏差についても良好になることはなく、前章の SOM による予測における全てのパラメータを使用したときの分子構造パラメータを圧縮しない場合の予測精度と、PCA を用いて分子構造パラメータを圧縮した場合の予測精度の比較結果と同じ傾向が確認できた。図 5-19 に示すのは OR67a, OR65a の応答に対して、一例として、使用する主成分个数(合成分子構造パラメータ个数)が 18 個の時の交差検定の実測値と予測値において、検討した PCA で圧縮して得た主成分に変数選択法を使用した SOM による ORN 応答予測の交差検定結果である。図 5-19(a) は OR67a に対して手法を適用したときの誤差のヒストグラム、図 5-19(b)はその実測値を横軸

に予測値を縦軸にとったプロット、図 5-19(c) は OR65a に対して手法を適用したときの誤差のヒストグラム、図 5-19(d)はその実測値を横軸に予測値を縦軸にとったプロットを表す。この図を見てもわかるとおり、予測精度が不十分であったため、それ以上の検討は中止した。

PCA で次元圧縮をし、そこで得られた合成変数を入力して、SOM 学習させる手法に関する検討の結果としては、精度は改善されなかった。PCA で合成変数化すると、本来有用であったパラメータが不要なパラメータと合成されてしまうのではないかと考察している。

### 5.7 蛍光プレートリーダーを用いた昆虫の嗅覚受容体応答測定の応答強度予測

現在神崎研究室において図 5-20 に示すような昆虫のイオトロピックレセプター遺伝子を蛍光タンパク質遺伝子と共に Sf21 細胞へ導入した、バイオセンサーとして利用可能なヘテロ遺伝子発現細胞の作製が報告されている [061]。神崎研究室では、カビ臭の中に含まれる 1-octen-3-ol に最も強く応答する OR13a を蛍光母細胞へ導入し、安定株種を作成する事に成功し、この細胞株種の香気への応答を取得した。本検討では、DoOR データベースにまとめられている OR13a 及び、神崎研究室で安定株種化された蛍光細胞の応答の応答強度予測について検討した。

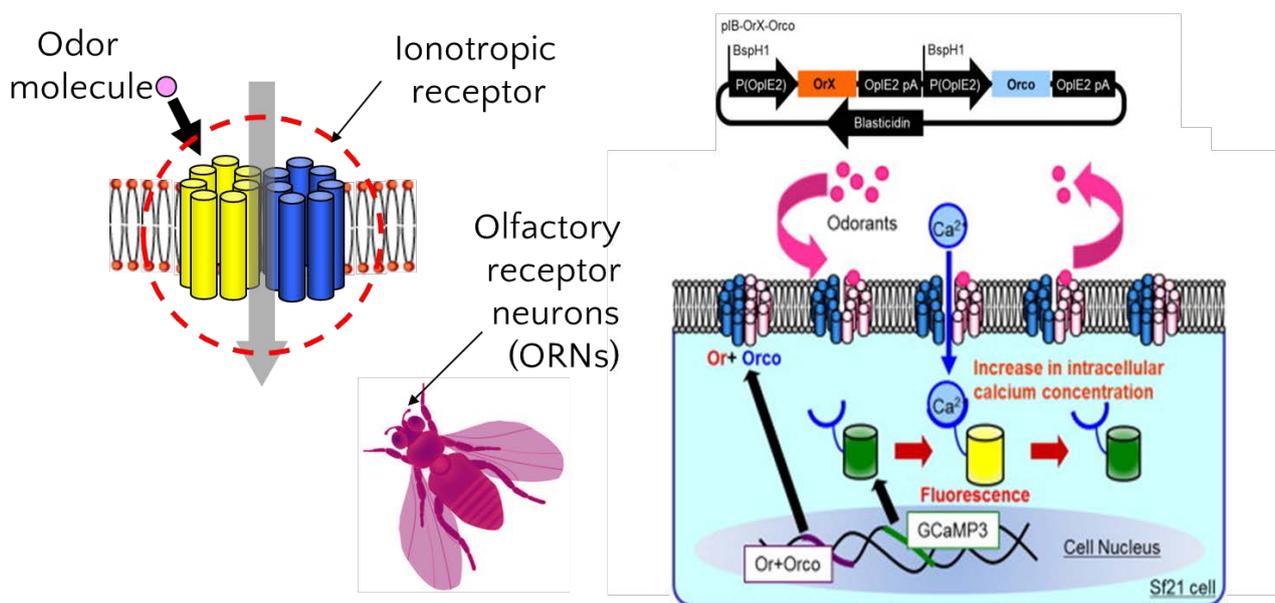


図 5-20. 昆虫の OR 遺伝子をヘテロ発現させて作成されたバイオセンサー細胞 [061]

嗅覚受容体 OR13a を導入した嗅覚受容体バイオセンサーが、東大神崎研究室で作成され、自動的に試薬の分注と励起光照射による蛍光測定が大規模に可能となる装置、蛍光プレートリーダー(以下 FDSS, Functional Drug Screening System [063])を用いてデータ取得が行われた。応答特性データは同研究グループの HP で ORN からの電気生理等で取得した応答値がまとめてある。図 5-21 に示すような OR13a の測定結果が参照できる。縦に応答取得された香気物質が並んでいる。香気物質は興奮性応答の大きいものから、抑制性応答の大きいもの、とい

う順序に並べてあり、棒グラフの長さとし色調はその大きさを示す(棒グラフの色調が赤いほど大きい興奮性の応答を示す)。

図 5-21. HP にて参照できる OR13a 応答特性データ [091]

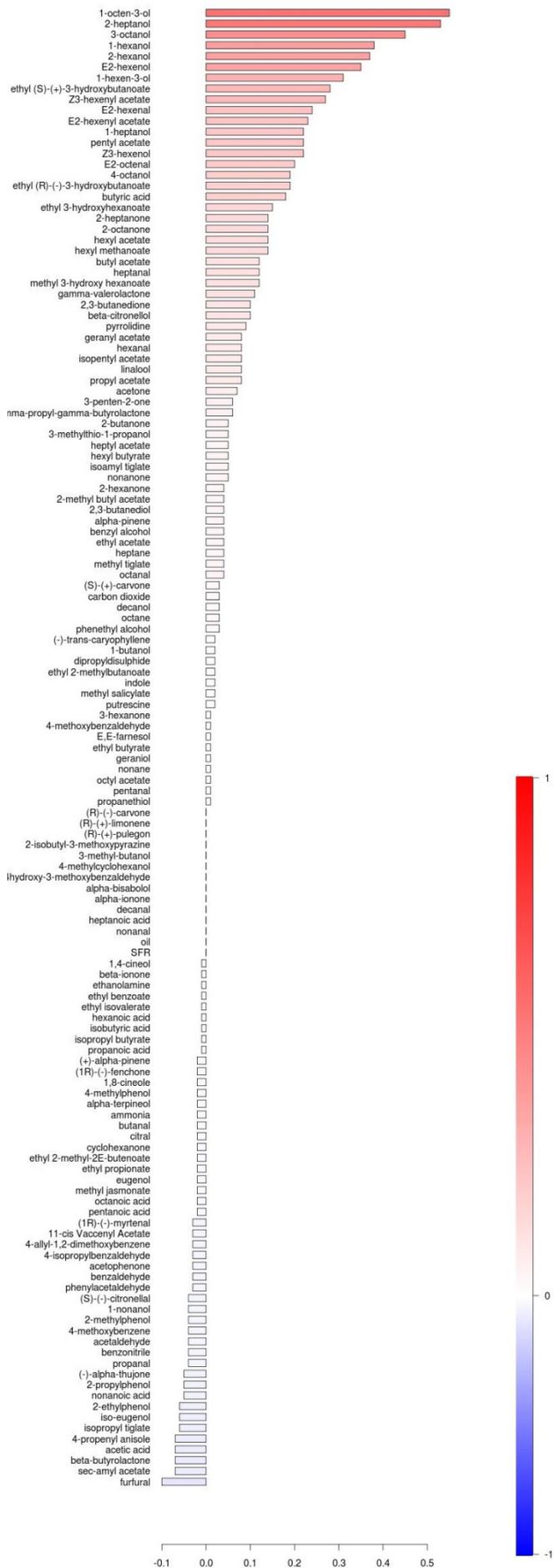


表 5-04. 応答測定された香気分子と既報の DoOR データベースで報告された香気物質の種類; (黒で名称を示した化合物は両試験方法で応答を調査されている、青で名称を示した化合物は FDSS で応答を調査されている、赤で名称を示した化合物は DoOR で応答を参照できる)

1-Butanol	6-Methyl-5-heptenone	isobutyric acid	methyl tiglate
1-Hexanol	4-Methyl-3-penten-2-one	E2-hexenol	ethyl 2-methyl-2E-butenate
Cyclohexanone	5-Nonanone	2,3-butanediol	isopropyl tiglate
Benzaldehyde	2,3-Dimethyl-2-cyclopenten-1-one	decanol	ethyl isovalerate
Heptanal	2-Methylcyclopentanone	1-heptanol	heptyl acetate
Decanal	3-Methylcyclopentanone	3-octanol	Z3-hexenyl acetate
Octanal	2-Methylbutyraldehyde	4-octanol	hexyl methanoate
Nonanal	2-Nonenal	2-heptanol	acetone
Acetic acid	2-Methylbutyric acid (racemic)	2-hexanol	2-butanone
Butyric acid	Isovaleric acid	1-hexen-3-ol	2,3-butanedione
Heptanoic acid	2-Methylheptanoic acid	acetaldehyde	3-penten-2-one
Octanoic acid	2-Methylhexanoic acid	propanal	2-hexanone
Hexanoic acid	Dimethyl sulfide	butanal	3-hexanone
Eugenol	Eugenol Acetate	pentanal	cyclohexanone
(-)- $\beta$ -caryophyllene	2,3,5-Trimethylpyrazine	E2-octenal	2-octanone
Propyl acetate	7-Octenoic Acid	putrescine	nonanone
Ethyl acetate	2-Butanone	ammonia	2-isobutyl-3-methoxypyrazine
Isoamyl acetate	2-Acetoxy-3-butanone	ethanolamine	pyrrolidine
Amyl acetate	1-Dodecanol	phenethyl alcohol	beta-butyrolactone
Furfural	VUAA1 ChemSpider ID: 1105882	benzyl alcohol	gamma-valerolactone
Indole	2-Propanol	methyl salicylate	gamma-propyl-gamma-butyrolactone
( $\pm$ )-Citronellal	2-Nonanone	ethyl benzoate	heptane
Geraniol	2-Furoic acid	phenylacetaldehyde	octane
1-Octen-3-ol	Adipic acid	acetophenone	nonane
Linalool	Pyridine	2-methylphenol	carbon dioxide
2-Heptanone	2,5-DimethylPyrazine	4-propenyl anisole	11-cis Vaccenyl Acetate
E2-Hexanal	Pyrazine	4-methoxybenzaldehyde	3-methylthio-1-propanol
1-Nonanol	2-Ethyl-3,(5 or 6)-dimethylpyrazine	4-isopropylbenzaldehyde	dipropyldisulphide
3-Methyl-1-butanol	1-Octanol	4-methoxybenzene	propanethiol
(1R)-(-)-Fenchone	2-Acetylthiophene	4-allyl-1,2-dimethoxybenzene	alpha-pinene
(R)-(-)-Carvone	m-Cresol	iso-eugenol	geranyl acetate
(S)-(+)-Carvone	(+)-Fenchone	4-methylphenol	alpha-terpineol
Valeric acid	Cadaverine	benzotrile	beta-citronellol
Ethyl 3-hydroxybutyrate	$\gamma$ -Decalactone	methyl jasmonate	1,8-cineole
4-Methylcyclonexanol (E/Z) mixture	$\sigma$ -Decalactone	2-ethylphenol	1,4-cineol
Ethyl propionate	Linalyl acetate	4hydroxy-3-methoxybenzaldehyde	citral
Cis-3-Hexen-1-ol	Ethyl hexanoate	butyl acetate	(R)-(+)-limonene
2-Propylphenol	Methyl octanoate	hexyl acetate	(1R)-(-)-myrtenal
Pentyl acetate	1-Chlorododecane	E2-hexenyl acetate	(+)-alpha-pinene
Hexanal	N-Methyldiethanolamine	ethyl butyrate	(R)-(+)-pulegon
1-Propanol	Ethyl formate	hexyl butyrate	(-)-alpha-thujone
1-Pentanol	Methyl propionate	octyl acetate	alpha-bisabolol
2-Ethylhexanol	4,5-Dimethylthiazole	2-methyl butyl acetate	E,E-farnesol
2-Methyl-1-butanol (racemic)	Cis-2-Hexen-1-ol	sec-amyl acetate	alpha-ionone
3-Methyl-2-buten-1-ol	1-Hepten-3-ol	ethyl 2-methylbutanoate	beta-ionone
3-Hexanol	Geosmin	ethyl 3-hydroxyhexanoate	
3,5,5-Trimethyl-1-hexanol	propanoic acid	isoamyl tiglate	
4-Methyl-2-pentanone	nonanoic acid	isopropyl butyrate	

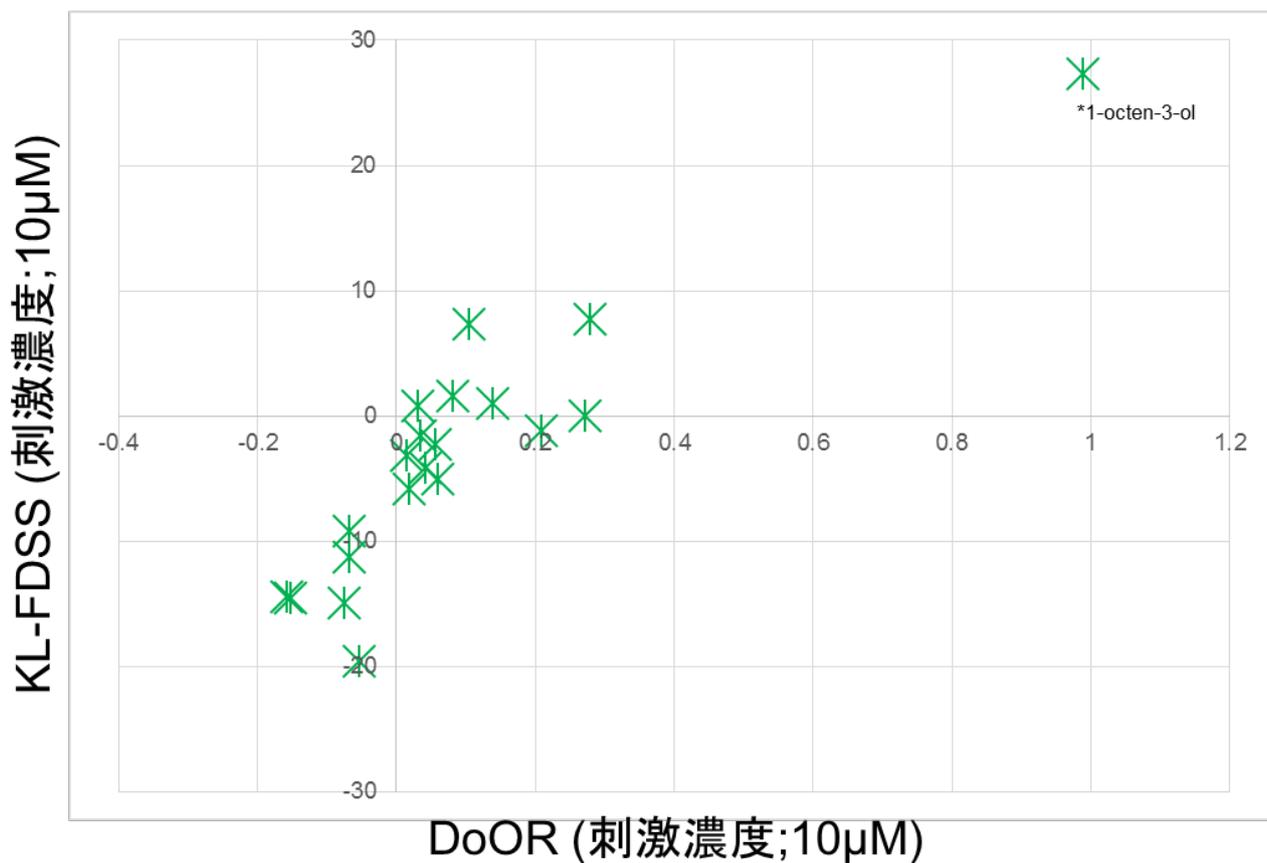
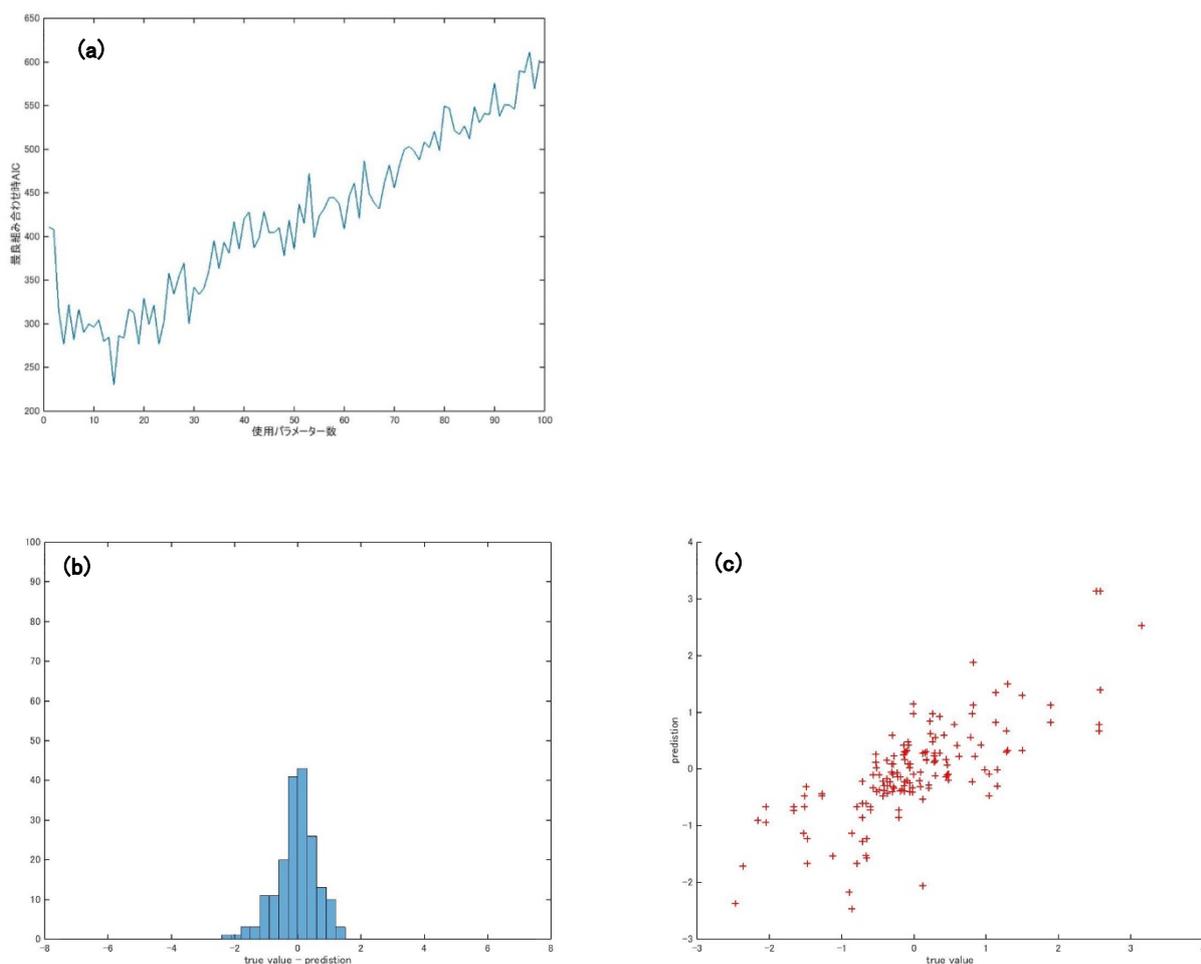


図 5-22. FDSS データベースと DoOR データベースで取得されている応答強度の線形性の確認

応答測定された香気分子の種類が既報の DoOR データベースで報告されたものとどれくらい重複していて測定結果がどれくらい線形性を持っているのかを確認した。表 5-04 に示したのは黒文字で示す化合物が両データベースで取得されている化合物、青文字で示す化合物が FDSS データベースで取得されている化合物、茶色文字で示す化合物が DoOR データベースで取得されている化合物である。

表 5-04 にて、黒文字で示す化合物が両データベースで取得されている化合物の応答強度を参照し、FDSS データベースと DoOR データベースで取得されている応答強度の線形性の確認をおこなった。横軸に FDSS データベースで取得されている化合物への応答強度(数値は蛍光強度変化量を標準刺激物質 VUAA を 100 として規格化)、縦軸に DoOR データベースで取得されている化合物への応答強度(数値は最大変化を示す香気物質による応答を 1 として規格化)、をプロットしたものを図 5-22 に示す。高い選択性を示すはずの香気物質に対する応答も中庸な応答を示すはずの香気物質、ほとんど応答しないはずの香気物質への応答がある程度の線形性を確保できていることが確認できた。



**図 5-23. FDSS データベース OR13a の応答強度の予測 ; (a)パラメータ数と交差検定における AIC の関係、(b)誤差のヒストグラム、(c)実測値と予測値の比較(両値オートスケーリング後)**

VUAA を除く全香気(化合物種 93 種類)に対するデータに対して条件数抽出した計算を反映して使用するパラメータは 100 種類、学習させる SOM のサイズが  $6 \times 6$  の条件で、パラメータを漸減する手法を適用した。これまでの検討と同じように、蛍光プレートリーダーを用いた昆虫の嗅覚受容体応答測定の結果をもとに、最適なパラメータの選定による予測手法が適用できることが確認できた。

その結果を図 5-23 にまとめる。パラメータ数が 14 個の時に、交差検定の AIC 評価が最小となった。パラメータを漸減する過程での AIC 値は図 5-23(a)のようにプロットされる。その条件における交差検定結果の実測値からの誤差をヒストグラムの形であらわすと図 5-23(b)のようになる。図 5-23(c)に示すのは交差検定の結果を横軸に実際の値、縦軸に予測された応答強度でプロットしたものである。これら図にて確認したところ、予測精度は比較的高かった。

## 5.8 SOMによる応答予測における3値化検討

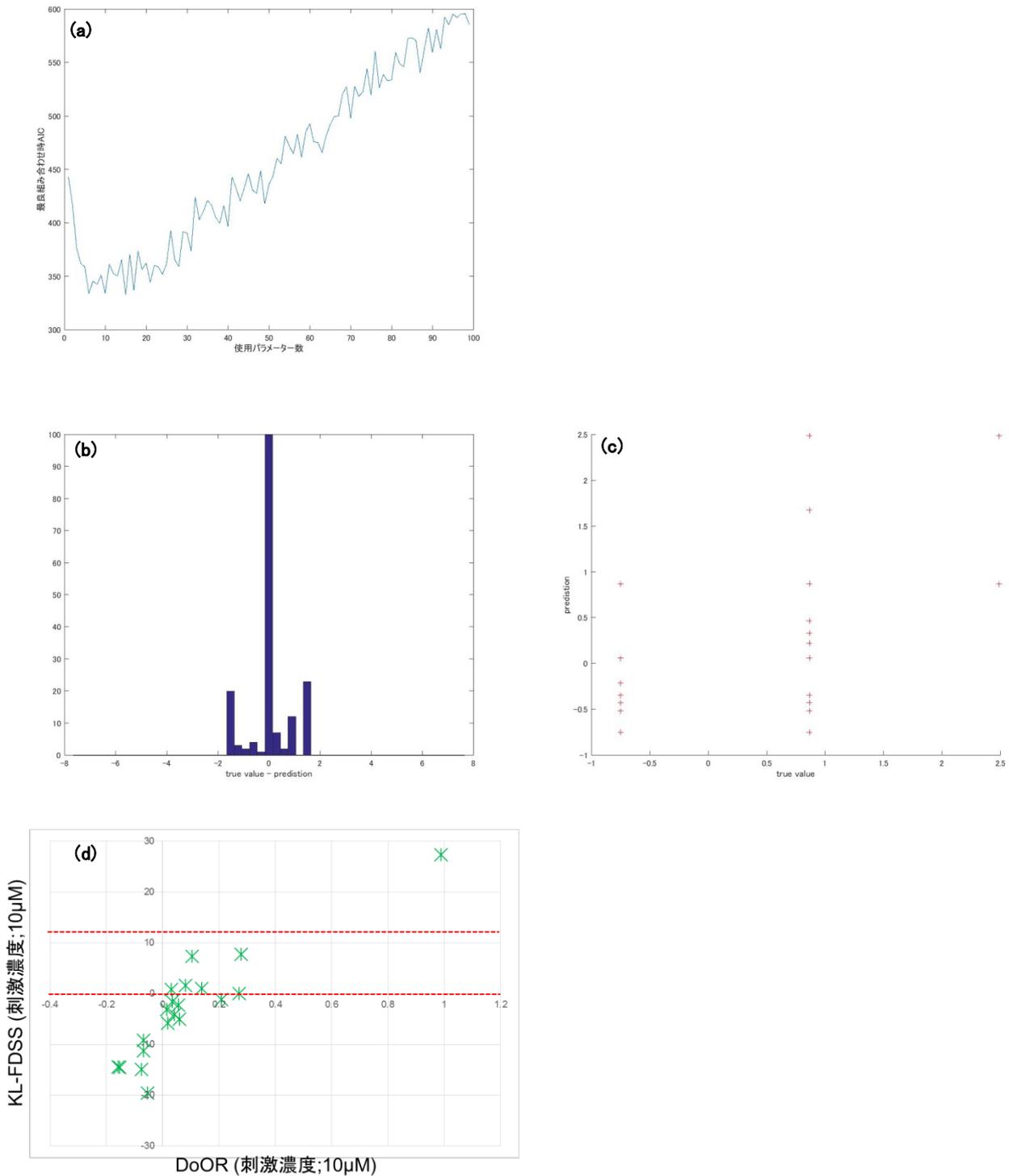


図 5-24. ORN 応答値の 3 値化処理の OR13a の応答強度の予測(a)パラメータ数と AIC の関係、(b)誤差のヒストグラム、(c)実測値と予測値の比較、(d)用いた閾値

予測精度の交差検定結果に対して再度閾値を設けて 3 値化をすることで、3 値化された応答強度予測が可能であると考えた。3 値化をする際に用いた閾値は図 5-24(d)に赤の破線にて示す。この閾値を用いて応答値を-1,0,1

としてこれまでと同じ手法を用いて、条件数+AIC 準拠パラメータ漸減探索したところ以下の結果を得た。

VUAA を除く全香気(化合物種 93 種類)に対するデータに対して条件数抽出した計算を反映して使用するパラメータは 100 種類、学習させる SOM のサイズが  $6 \times 6$  の条件で、パラメータを漸減する手法を適用した。これまでの検討と同じように、蛍光プレートリーダーを用いた昆虫の嗅覚受容体応答測定の結果をもとに、最適なパラメータの選定による予測手法が適用できることが確認できた。

その結果を図 5-24 にまとめる。パラメータ数が 15 個の時に、交差検定の AIC 評価が最小となった。パラメータを漸減する過程での AIC 値は図 5-24(a)のようにプロットされる。その条件における交差検定結果の実測値からの誤差をヒストグラムの形であらわすと図 5-24(b)のようになる。図 5-24(c)に示すのは交差検定の結果を横軸に実際の値、縦軸に予測された応答強度でプロットしたものである(実測値、予測値共に応答強度はオートスケーリングされている)。これら図にて確認したところ、図 5-23 に示した未 3 値化における予測に近い予測精度が得られた。

AIC 最小値を出した時の使用パラメータ個数は大体同じくらいになった。閾値を微調整することで実用的な予測精度を出せると考える。その反面、神崎研究室では、1-octen-3-ol、1-hepten-3-ol(図 5-22 にはプロットされていない)についてのみ有意な応答が得られているという解析となっているため、今後はバイナリ値化時の予測における閾値をどのように設定するかに関して検討する際には注意を怠ってはならない。

## 5.9 まとめ

香気分子の化合物構造を元にした嗅覚受容体予測手法の検討した。

香気分子の構造を Talete DRAGON でパラメータ化し、それを元に SOM をトレーニングすると分子情報を元にしたクラスタ化ができる。この時に嗅覚受容特性を合わせてトレーニング要素として配置すると、トレーニングされた SOM の中にも受容特性情報が合わせてマッピングされる、最近傍ベクトル参照してやることで、応答未知の化合物に対しても応答を予測することが出来る。

SOM 学習においては、距離尺度は線形なものが適用されるため、各パラメータごとの規格化は重要なポイントになる。当初はこの規格化条件も学習によって決定するようなモデルへと発展させることも考えたが、そのアイデアを実装することはなかった。距離に関して検討したのは、全てのパラメータを規格化なしで利用した場合と、平均 0、標準偏差 1 に線形変換した場合の二つだけとした。結果としては、この線形変換は必須である。また SOM トレーニング時における距離尺度は一般的にはユークリッド距離となるが、IS 距離などの特殊な距離尺度を導入する事で、より特徴抽出された SOM 学習が期待されるが、現状未検討である。

化合物の分子構造パラメータは幾分冗長なパラメータ空間を形成している物と思われる。も本章の検討では、条件数を用いて、冗長性を排除した。その後、予測性能を見ながら、利用するパラメータを漸減しながら最適予測パラメータ組み合わせを探索するという 2 段階のパラメータ削減手法を提案した。

OR65a, OR13a のデータへ応答予測法を適用したところ、OR13a のデータについて比較的良好な予測が得られた。

神崎研究室で新しい蛍光プレートリーダーを用いてOR13aの応答データが取得された。このデータとDoORデータベースで重複している香気分子への応答測定結果を参比較して、高い選択性を示すはずの香気物質に対する応答も中庸な応答を示すはずの香気物質、ほとんど応答しないはずの香気物質への応答がある程度の線形性を確保できていることが確認できた。このOR13aデータに関して、これまで検討され確立した、条件数検討をはじめとするパラメータ削減手法を適用し、受容応答予測を実施した。SOMによる予測では、かなり近い予測精度を得ることができた。

## 第六章 香気分子構造情報のオートエンコーダーによる次元圧縮と 多層パーセプトロンによる嗅覚受容体の応答強度予測手法

化合物の分子構造パラメータは幾分冗長なパラメータ空間を形成していると思われる。もっとも最初に行った検討では、条件数を用いて、冗長性を排除しようとしたのだが、空間の圧縮方法を適用し、合成変数化して、予測をする手法も有用ではないかと当初より考えられていた。次元圧縮し、合成パラメータを生成する手法としては、PCA(主成分分析)やNMFのような、線形変換ベースで直交性を保証できる可能性が高いものや、非線形性の高いオートエンコーダーのような手法がある。本章では、オートエンコーダーを検討した。

### 6.1 オートエンコーダー(砂時計型ニューラルネットワーク)

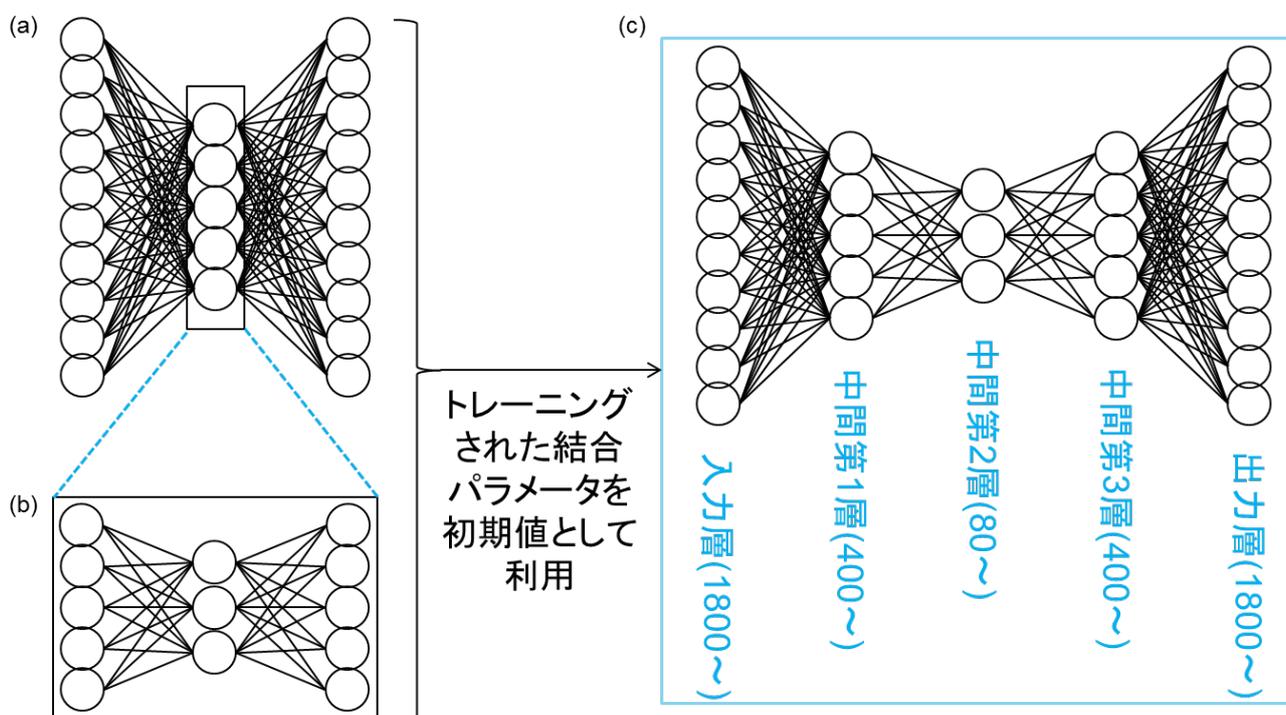


図 6-01. オートエンコーダーのバックプロパゲーション学習手順；(a)第一段階の学習、(b)第二段階の学習、(c)第三段階の学習；二種類のトレーニングで得られた結合パラメータをもとにトレーニング

オートエンコーダは多層ニューラルネットにおいて、入力層と出力層に同じデータを用いて教師あり学習させたものである。データ本来の情報の情報量を保ちつつ次元圧縮を行う。出力が入力を再現するように学習されるため、得られた中間層が特徴抽出された情報となる

香気分子の構造をもとにした嗅覚受容体(ORN)の応答予測における構造パラメータの次元圧縮のために、バックプロパゲーション学習させて得られたオートエンコーダー(砂時計型ニューラルネットワーク)の利用を検討した。

図 6-01. にオートエンコーダーのバックプロパゲーション学習手順を示した。学習は以下の三段階からなる。

- 第一段階の学習; 図 6-01(a)のように入力層と出力層に同じ教師データを与えて層間の結合パラメータをトレーニングする
- 第二段階の学習; 図 6-01(b)のようにトレーニングされた結合パラメータを用いて入力層を転写し、より少ないユニット数の中間ユニット層を経由する結合パラメータをトレーニングする。
- 第三段階の学習; 図 6-01(c)のように二種類のトレーニングで得られた結合パラメータをもとに、3つの中間層を経る各結合パラメータをトレーニングする。この中間ユニット層には次元圧縮された写像が表れる。

オートエンコーダーを有効に働かせるためには、さまざまなパラメータを設定する事が必要であるが、その条件の検討は手間と時間を要することが一般的である。

## 6.2 分子構造パラメータ空間の次元圧縮

今回行うオートエンコーダーの検討は、図 6-02 に示すような香気物質の分子構造パラメータの次元圧縮と言える。オートエンコーダーにより得られた中間層を利用する機械学習の手法は広く知られている。

本章では再び神崎研究室で蛍光プレートリーダーを用いて取得された OR13a の応答データの応答予測の検討を実施した。応答取得されている香気物質に関する分子構造パラメータ空間について、香気分子の構造を Taletе DRAGON を利用してパラメータ化すると、およそ 5000 次元のパラメータが得られるが、それら種類のうち半分ほどに、すべての分子種に対して Taletе DRAGON 出力が一定値または NaN を含んでいたので、そのような分子パラメータを除外し、1800 次元のパラメータを得た。オートエンコーダーによる圧縮を行い、二種類の予測手法を検証した。一つは合成パラメータの漸減プロセスを導入した、5 章にて説明した SOM による予測である。もう一つは中間層を一層挟むニューラルネットワークへのバックプロパゲーション学習による予測システムである。

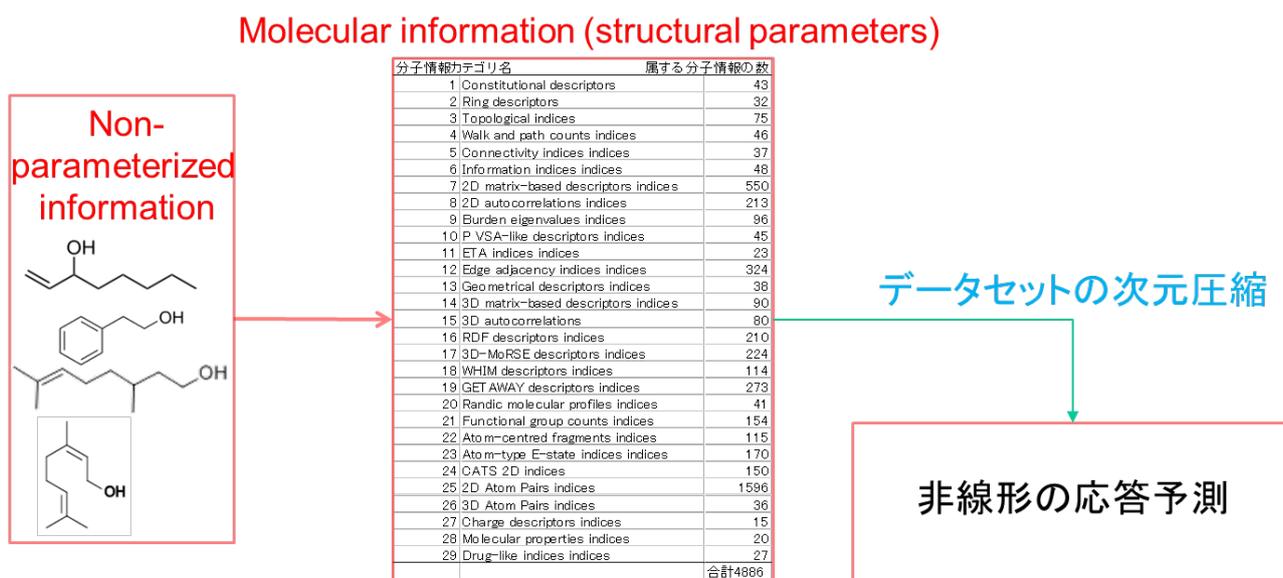


図 6-02. オートエンコーダーを用いた香気分子の構造パラメータの次元圧縮

オートエンコーダーを用いた空間圧縮は、様々な学習係数を検討する必要がある。学習係数のうち、各層におけるユニット数を大まかに変化させて、それらのおおよその最適値を探索した。次に固定した各層のユニット数に対して、各段階の学習係数、ハイパーパラメータや学習係数と学習に伴う係数変更のスケジュールを、適切な学習ができるように探索した。オートエンコーダーによるパラメータ圧縮では、いくつかのユニット数それぞれを求め、いずれのユニット数を選択しても予測手法を検証できるようにオートエンコーダーによる次元圧縮を計算した。

### 6.3 香気分子の分子構造パラメータへのオートエンコーダー(砂時計型ニューラルネットワークの適用)

#### 6.3.1 データについて

表 6-01. オートエンコーダーで次元圧縮する対象として選んだ香気分子

1	1-Propanol	32	2-Methylheptanoic acid	63	Pyrazine
2	1-Butanol	33	2-Methylhexanoic acid	64	2-Ethyl-3,(5 or 6)-dimethylpyrazine
3	1-Pentanol	34	Dimethyl sulfide	65	1-Octanol
4	2-Ethylhexanol	35	Eugenol	66	2-Acetylthiophene
5	1-Hexanol	36	Eugenol Acetate	67	m-Cresol
6	2-Methyl-1-butanol	37	(-)- $\beta$ -caryophyllene	68	(+)-Fenchone
7	3-Methyl-2-buten-1-ol	38	2,3,5-Trimethylpyrazine	69	(1R)-(-)-Fenchone
8	3-Hexanol	39	7-Octenoic Acid	70	(R)-(-)-Carvone
9	3,5,5-Trimethyl-1-hexanol	40	2-Butanone	71	(S)-(+)-Carvone
10	4-Methyl-2-pentanone	41	2-Acetoxy-3-butanone	72	Cadaverine
11	Cyclohexanone	42	Propyl acetate	73	Valeric acid
12	6-Methyl-5-heptenone	43	Ethyl acetate	74	$\gamma$ -Decalactone
13	4-Methyl-3-penten-2-one	44	1-Dodecanol	75	$\sigma$ -Decalactone
14	5-Nonanone	45	Isoamyl acetate	76	Linalyl acetate
15	2,3-Dimethyl-2-cyclopenten-1-one	46	Amyl acetate	77	Ethyl 3-hydroxybutyrate
16	2-Methylcyclopentanone	47	Furfural	78	Ethyl hexanoate
17	3-Methylcyclopentanone	48	Indole	79	Methyl octanoate
18	2-Methylbutyraldehyde	49	( $\pm$ )-Citronellal	80	1-Chlorododecane
19	Benzaldehyde	50	Geraniol	81	N-Methyldiethanolamine
20	Heptanal	51	1-Octen-3-ol	82	4-Methylcyclonhexanol
21	Decanal	52	Linalool	83	Ethyl formate
22	Octanal	53	2-Heptanone	84	Ethyl propionate
23	Nonanal	54	E2-Hexanal	85	Methyl propionate
24	2-Nonenal	55	2-Propanol	86	4,5-Dimethylthiazole
25	Acetic acid	56	1-Nonanol	87	Cis-2-Hexen-1-ol
26	Butyric acid	57	3-Methyl-1-butanol	88	Cis-3-Hexen-1-ol
27	Heptanoic acid	58	2-Nonanone	89	1-Hepten-3-ol
28	2-Methylbutyric acid	59	2-Furoic acid	90	2-Propylphenol
29	Octanoic acid	60	Adipic acid	91	Pentyl acetate
30	Hexanoic acid	61	Pyridine	92	Geosmin
31	Isovaleric acid	62	2,5-DimethylPyrazine	93	Hexanal

表 6-01. に本検討で使用したオートエンコーダーで次元圧縮する対象として選んだ香気分子を示す。この化合物群は嗅覚受容体(OR13a)発現細胞の蛍光観測にて用いられた 93 種類の香気物質であり、蛍光プレートリーダーを用いた応答測定の結果(詳細は 5.7 節を参照)をもとに、オートエンコーダー圧縮条件を検討し、二種類の予測を検証した。

### 6.3.2 オートエンコーダーのバックプロパゲーション学習手順

誤差関数は、次の式で表せる。

$$E(w) = \frac{1}{2} \sum_{i=1}^N \|y(\text{input}_i, w) - \text{train}_i\|^2 \quad (\text{式 6.1})$$

多層ニューラルネットワークにおける活性化関数は標準シグモイド関数

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (\text{式 6.2})$$

を用いた。今回、オートエンコーダへの学習は確率的勾配降下法を用いた。これはバックプロパゲーション学習におけるオンライン学習アルゴリズムの一種であり、ランダムな順番でサンプル選び、一サンプルずつデータの微分値(勾配)を計算してパラメータを

$$W_{t+1} = W_t + \eta \frac{\partial E}{\partial x} W_t \quad (\text{式 6.3})$$

のように繰り返し更新する学習方法である。今回、これを拡張した

$$W_{t+1} = W_t + \eta \frac{\partial E}{\partial x} W_t - \alpha W_{t-1} + \varepsilon \quad (\text{式 6.4})$$

を用いた。 $\eta$ は学習係数である。ランダムに並べ替えた教師信号による学習を1サイクル行うたびに減少させた。 $\alpha W_{t-1}$ は慣性項と呼ばれ、収束速度を向上させられるといわれている。ごく小さい値 $\varepsilon$ は学習時に掛けられるゆさぶりである。今回、一般的な正則化項を考慮した誤差関数、

$$E(w) = \frac{1}{2} \sum_{i=1}^N \|y(\text{input}_i, w) - \text{train}_i\|^2 + \lambda \|W\|^q \quad (\text{式 6.5})$$

を採用した。 $q=1$ のときは1ノルムによる正則化であり、L1正則化またはLASSO正則化と呼ばれる。 $\lambda$ はハイパーパラメータと呼ばれ、正則化の強さに関係している。今回の検討においてはLASSOタイプの正則化法を検討した。

図6-01. にオートエンコーダーのバックプロパゲーション学習手順を示したように①得られた約1800次元のパラメータを400次元に圧縮し、中間層1層のオートエンコーダーを得た。②次にこの圧縮された400次元のパラメータを75次元に圧縮し、中間層1層のオートエンコーダーを得た。③最後に一層目、三層目が400次元、二層目が75次元のオートエンコーダーをバックプロパゲーション学習し、中間3層のオートエンコーダーを得ることができる。本検討では、以降特に断りのない限り、上述のような活性化関数とバックプロパゲーション学習を行った。

### 6.3.3 中間層のユニット数

バックプロパゲーション学習を効率的に行うために、また得られたオートエンコーダーを有効に働かせるために、いくつかの条件を検討した。

まず中間層のユニット数を大まかに変化させ決定した。その結果、学習における残差と中間第一層及び中間第三層のユニット数に対する依存性は見いだせなかった。中間第二層のユニット数はバックプロパゲーション学習における収束の安定性、残差の大きさ、クロスバリデーションの結果に大きく影響するようである。今回のオートエ

ンコーダーによる次元圧縮では、まず応答予測における有効性を確認することを目標とした。そのため次元圧縮におけるユニット数を最適化することが必ずしも必要ないので、各層のユニット数の最適化は行っていない。以降6.3節では、中間第一層及び中間第三層のユニット数を400程度、中間第二層のユニット数は75程度としたオートエンコーダーの学習過程について説明をする。

### 6.3.4 バックプロパゲーション学習の方法

続いて、3段階あるバックプロパゲーション学習プロセスにおける各ステップにおいて、ハイパーパラメータと学習係数(ETA, EPSILON)の選択をした。交差検定のデータはランダムに20%選択し、残り80%を学習データとした。

各段階における学習の結果をもとに、学習回数に対する各段階の入出力誤差のプロット、学習データ、交差検定のデータそれぞれの実データ~出力値プロットを作成した。実データ~出力値プロットはすべてのパラメータの交差検定の結果を学習データ、交差検定のデータそれぞれまとめてプロットしてある(したがってプロット数はパラメータ数×サンプル数分ある)。今回、各段階での学習に使用した学習係数の条件において、良好にバックプロパゲーション学習することができることを確認した。以降に各段階における学習について説明をする。

**表 6-02. 実験にて使用した各パラメータ**

	ETA_init	ETA_bias / cycle	ALPHA	EPSILON	LASSO	HYPERPARAMETER(LASSO)	hidden_neuron_amount	max_iter
学習第1段階	0.4	0.98	0.3	0.0001	(on)	3.00E-09	400	3.20E+03
学習第2段階	0.4	0.98	0.3	0.0001	(on)	3.00E-07	80	3.20E+03
学習第3段階	0.2	0.98	0.3	0.0001	(on)	3.00E-10	-	3.20E+03

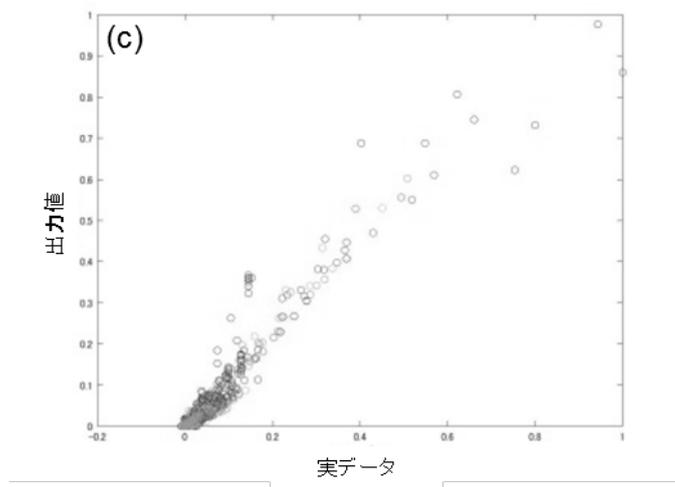
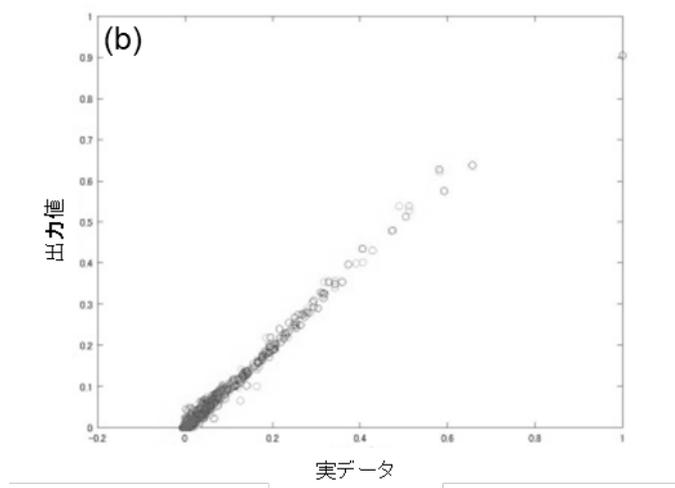
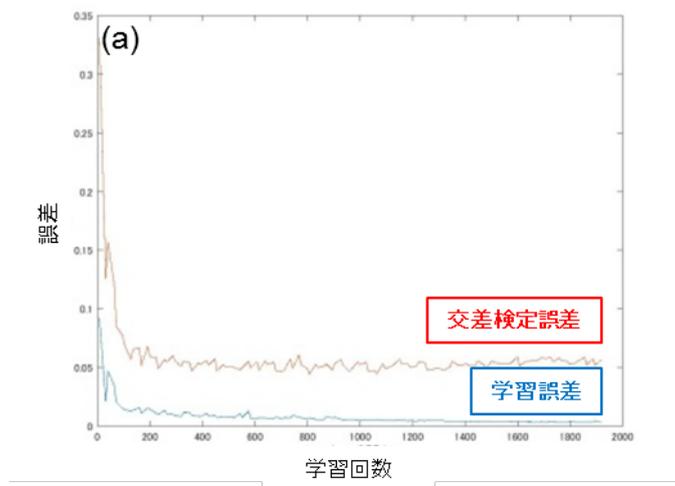
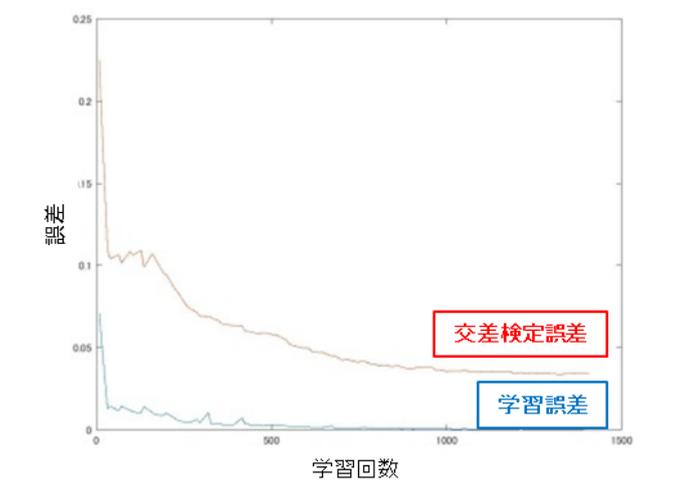


図 6-03. 第一段階目におけるバックプロパゲーション学習をした結果 ; (a) はオートエンコーダ学習時の誤差推移プロット、(b) は交差検定に基づく学習データ(80%)に対する実測値を横軸に予測値を縦軸にとったプロット、(c) 交差検定に基づく検定データ(残り 20%)に対する実測値を横軸に予測値を縦軸にとったプロット

図 6-03 に示す 3 つのグラフは第一段階目におけるバックプロパゲーション学習をした結果である。Hyper parameter 1 = 1.00E-4, Hyper parameter 2 = 3.00E-8 の条件でバックプロパゲーション学習を最大 3200 回(学習誤差の減少が止まったら学習を停止させたので本例では約 1900 回)行った結果、[1800-400-1800]の構造のオートエンコーダーが得られ、図 6-03a に示すように、トレーニングに伴う入出力値間残差およびクロスバリデーションデータにおける残差は減少した。学習終了時における学習データおよびクロスバリデーションデータの入力値と出力値のプロットは図 6-03b,c に示すように、ある程度良好な関係に収束していることが分かった。

第二段階目の学習では、Hyper parameter 1 = 1.00E-4, Hyper parameter 2 = 3.00E-7 の条件で最大 3200 回(学習誤差の減少が止まったら学習を停止させたので本例では約 1400 回)バックプロパゲーション学習させた。中間第一層及び第三層から中間第二層を計算するこの過程で、トレーニングに伴う入出力値間残差およびクロスバリデーションデータにおける残差は図 6-04 に示すように減少し、[400-80-400]の構造のオートエンコーダーにおいて良好な学習結果が得られたことがわかる。



**図 6-04. 第二段階目におけるバックプロパゲーション学習時の誤差推移プロット**

第三段階目においては第一段階、第二段階で得られた結合係数を初期値として学習最大 3200 回(学習誤差の減少が止まったら学習を停止させたので本例では約 600 回)を行った。[1800-400-80-400-1800]の構造のオートエンコーダーが得られ、トレーニングに伴う入出力値間残差およびクロスバリデーションデータにおける残差は図 6-04\_2 に示すように減少した。学習終了時における学習データおよびクロスバリデーションデータの入力値と出力値は図 6-04\_2 に示すように、ある程度良好な関係に収束し、良好な学習結果が得られたことがわかる。

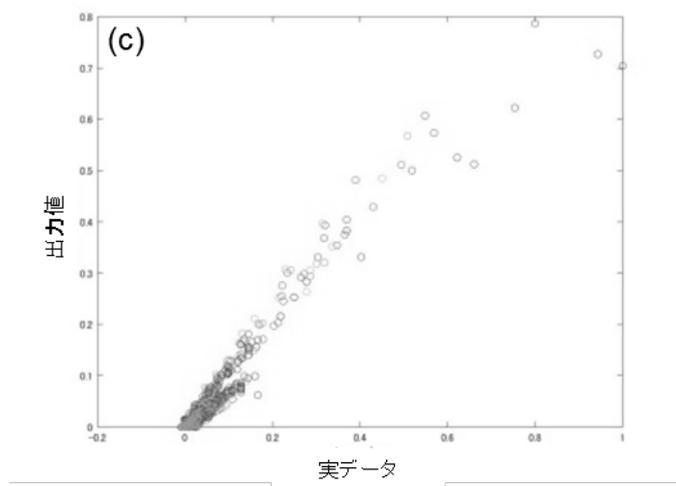
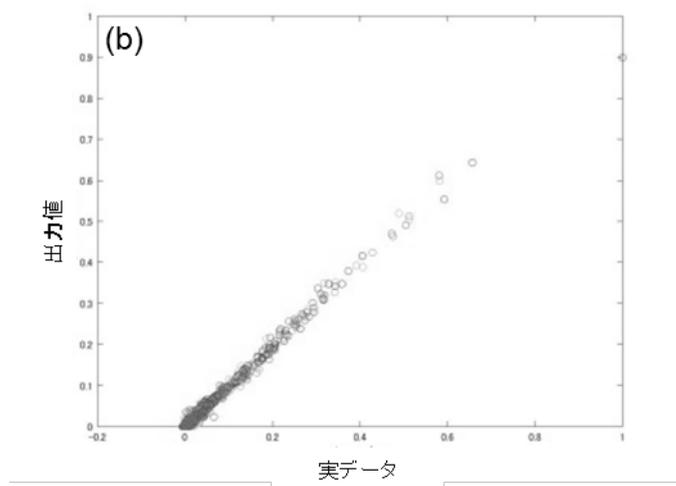
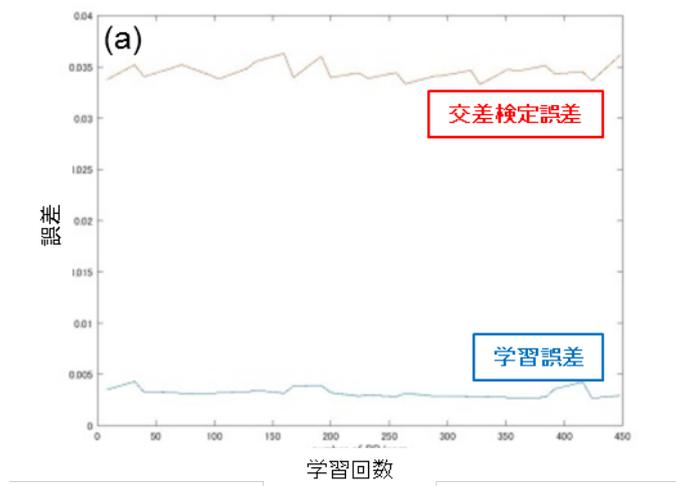
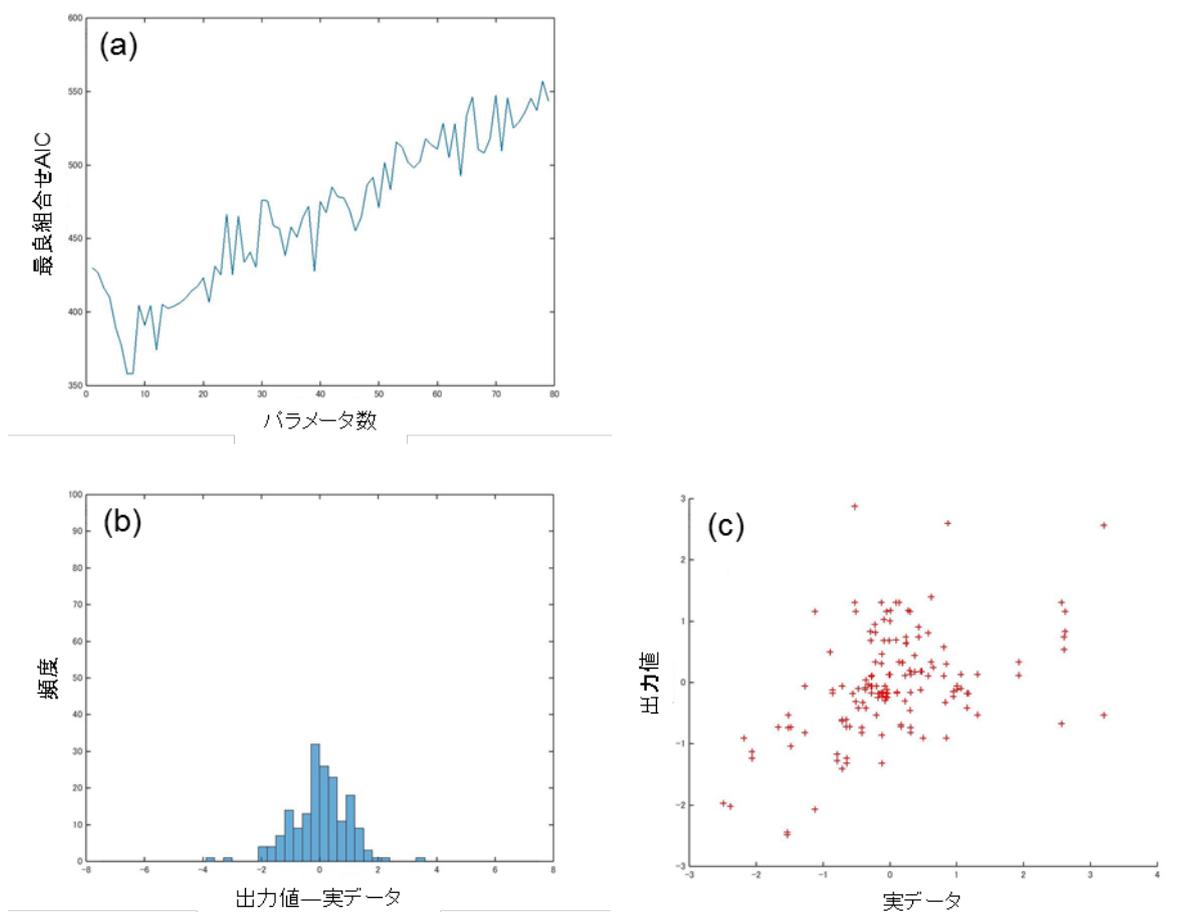


図 6-04\_2. 第三段階目における 5 層ニューラルネットワークへのバックプロパゲーション学習時の誤差推移プロット ; (a) はオートエンコーダ学習時の誤差推移プロット、(b) は交差検定に基づく学習データ(80%)に対する実測値を横軸に予測値を縦軸にとったプロット、(c) 交差検定に基づく検定データ(残り 20%)に対する実測値を横軸に予測値を縦軸にとったプロット

以上のようにオートエンコーダーによる次元圧縮の計算条件を求めた後、非線形な予測手法として、中間層を一層挟むニューラルネットワークへのバックプロパゲーション学習による予測、および最適なパラメータ組み合わせの探索を組み合わせた、分子情報を元にした SOM による予測について検証した。

## 6.4 SOM を用いた応答予測

### 6.4.1 オートエンコーダーによる香気化合物の分子パラメータを圧縮と SOM を用いた応答予測(93 化合物～OR13a)



**図 6-05. オートエンコーダー圧縮条件を元にした合成パラメータの漸減プロセスを伴うSOM による予測を行った結果; (a) は OR13a に対して手法を適用したときの AIC プロット、(b) は最も良い AIC の時のパラメータ組合せを適用したときの交差検定結果による誤差のヒストグラム、(c) はその実測値を横軸に予測値を縦軸にとったプロット**

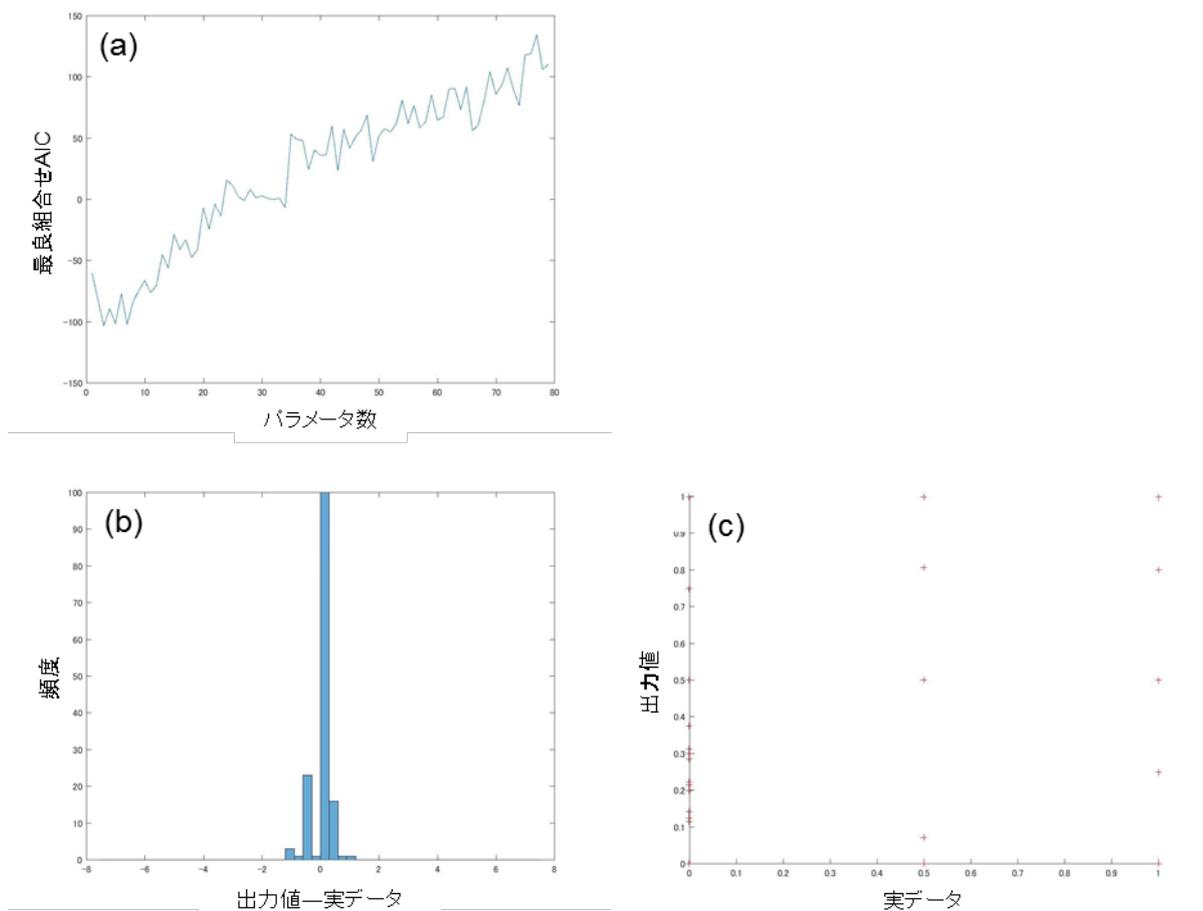
オートエンコーダーによる次元圧縮の計算条件を求めた後、その決定された計算条件を用いて、化合物の分子パラメータを圧縮した。圧縮された分子パラメータと応答測定結果をもとに、これまで検討されてきた SOM を用いた応答予測を実施した。この応答予測ではこれまで検討してきた条件数による分子パラメータの絞り込みは行わないものの、一段階ごとに予測結果が良くなかったパラメータを一つずつ減らすことで、最良の分子パラメータの組み合わせを探索した。

その決定された計算条件を用いて、神崎研究室にて取得されている OR13a について受容特性測定された化合物の分子パラメータを圧縮した。圧縮された分子パラメータと応答測定結果をもとに、これまで検討されてきた 6×6 のサイズの SOM を用いた応答予測を実施した(手法と各パラメータは 5.3 節にて説明をしたものを使用した)。この応答予測ではこれまで検討してきた条件数による分子パラメータの絞り込みは行わず、一段階ごとに予測結果が良くなかったパラメータを一つずつ減らすことで、最良の分子パラメータの組み合わせを探索した。オートエンコーダー圧縮条件を元にした合成パラメータの漸減プロセスを伴う SOM による予測を行った結果を図 6-05(a)-(c) に示す。合成分子パラメータが約 10 個の時に最良の予測となっていることがわかった。また図 6-05 (c)の結果から、これまでの SOM による予測に近い精度を持っていることが分かった。

#### 6.4.2 SOM を用いた応答予測における 3 値化の効果

SOM による応答予測における 3 値化検討を 5.8 節で検討したが、その効果をオートエンコーダーによる香気化合物の分子パラメータを圧縮と SOM を用いた応答予測でも検証した。

オートエンコーダーによる次元圧縮の計算条件を求めた後、神崎研究室にて取得されている測定された化合物の分子構造パラメータ空間の圧縮を決定された計算条件を用いて行った。圧縮された分子パラメータと応答測定結果をもとに、5.8 節にて検討した 3 値化の手法を用いて、SOM による応答予測を実施した(閾値、SOM 学習条件は 5.8 節に準じた)。



**図 6-06. ORN 応答値の 3 値化処理の検討; 下段が応答値を 3 値化し、同様の手法を適用した結果**

図 6-06.に ORN 応答値の 3 値化処理の検討の結果を示す。5.8 節と同様に閾値を設けて再 3 値化をすることで、同様の手法を適用した。AIC 最小値を出した時の使用パラメータ個数は大体同じくらいになった。5.8 節での考察とも重なるが、予測(この場合交差検定)結果に対して再度閾値を設けて再 3 値化をすることで、3 値化された応答強度予測が可能。閾値を微調整することで実用的な予測精度を出せると考える。



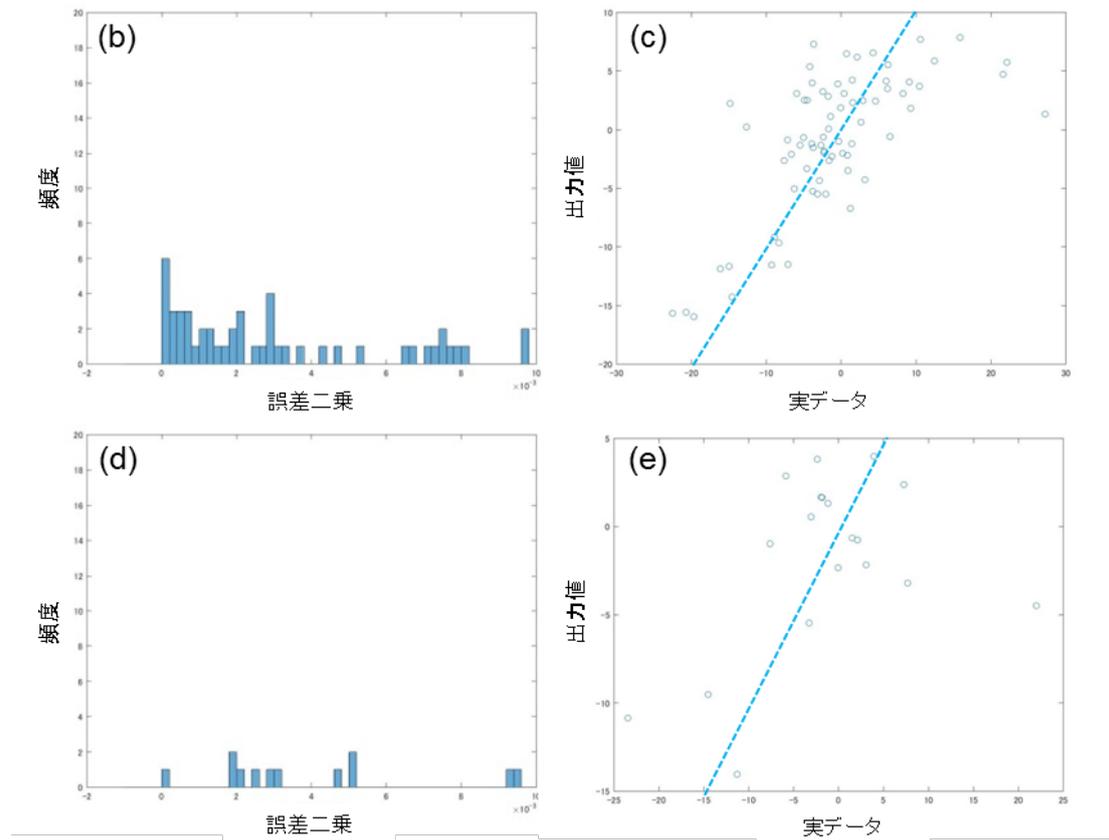
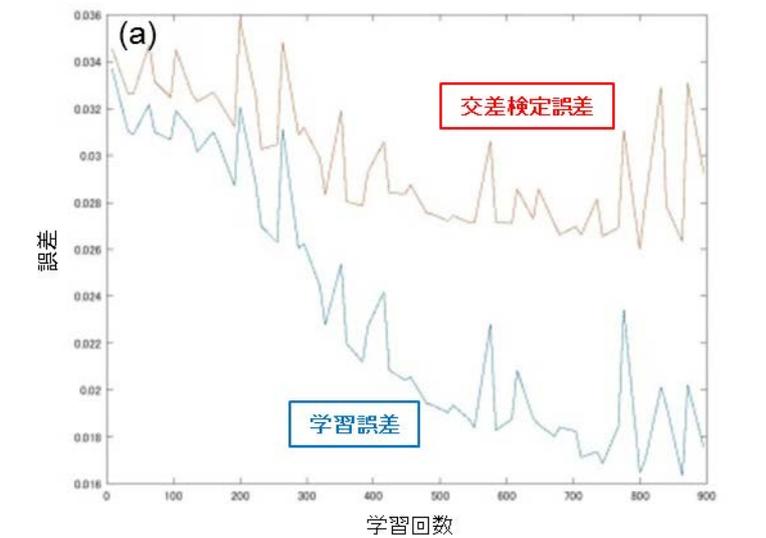


図 6-08. 80 次元まで圧縮した香気パラメータをもとにしたニューラルネットワークを用いた応答予測:(a)バックプロパゲーション学習過程におけるエラー推移、(b)トレーニングデータ(予測値-実測値)<sup>2</sup>ヒストグラム、(c) トレーニングデータ予測値-実測値プロット、(d)交差検定(予測値-実測値)<sup>2</sup>ヒストグラム、(e) 交差検定予測値-実測値プロット

### 6.5.2 オートエンコーダーによる全般的な香気化合物の分子パラメータを圧縮

さらに神崎研究室において応答を取得されていない化合物も含めた 1500 種類の香気化合物の構造パラメータに対してオートエンコーダーを用いた構造情報空間の圧縮を試みた。嗅覚受容体(OR13a)発現細胞の蛍光観測にて用いられた 93 種類の香気物質のパラメータ空間に適用に適用したのと同じ圧縮条件を用いた。

図 6-09. に 1500 種類の香気化合物の構造パラメータに対してオートエンコーダー[1800-40-80-400-1800]を使用し、80 次元まで、構造情報空間の圧縮結果を示す。

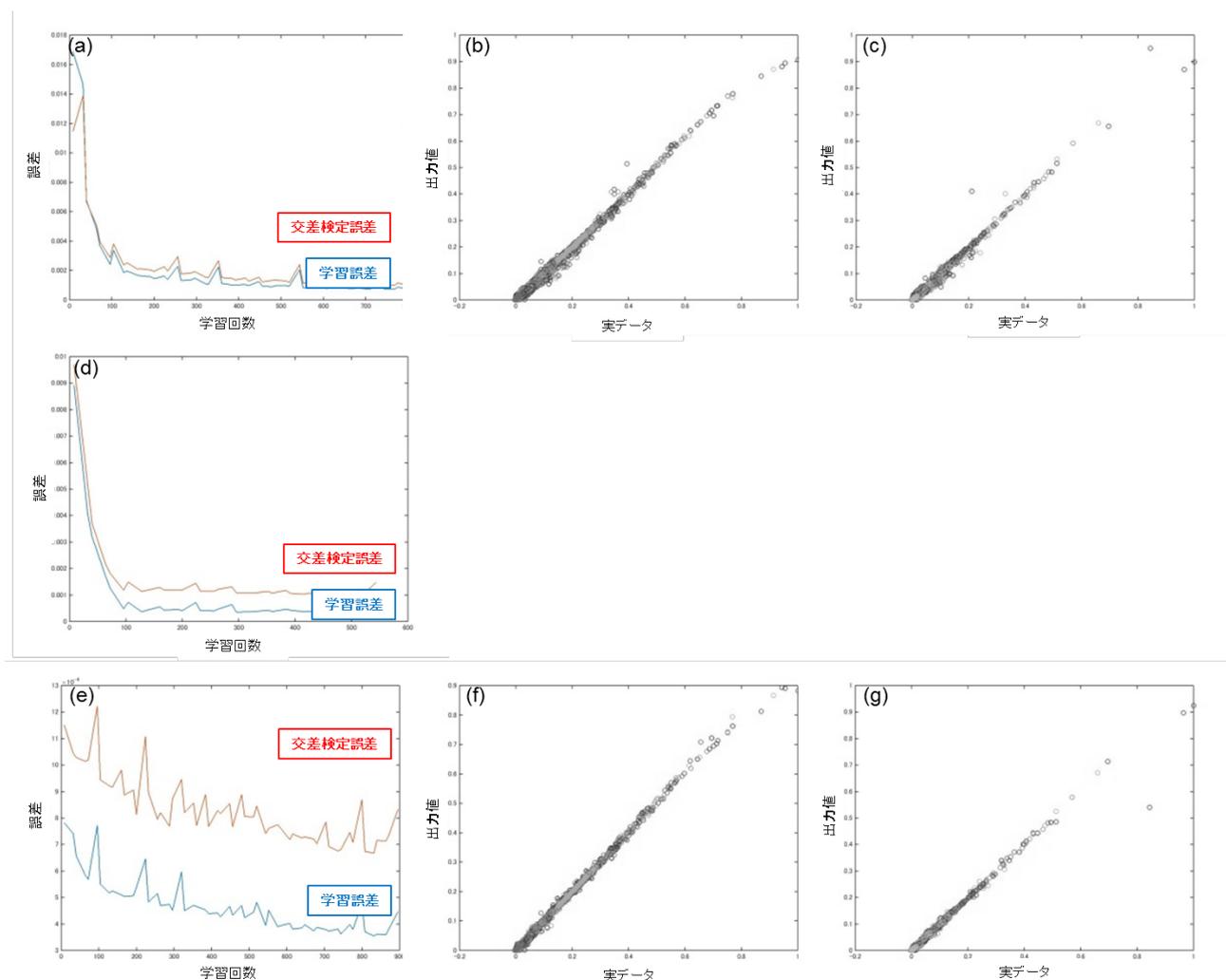


図6-09. 1500種類の香気化合物の構造パラメータに対してオートエンコーダーを用いた構造情報空間の圧縮 ; (a) は第一段目の学習におけるオートエンコーダ学習時の誤差推移プロット、(b) は第一段目の学習における交差検定に基づく学習データ(80%)に対する実測値を横軸に予測値を縦軸にとったプロット、(c) は第一段目の学習における交差検定に基づく検定データ(残り20%)に対する実測値を横軸に予測値を縦軸にとったプロット (d) は第二段目の学習におけるオートエンコーダ学習時の誤差推移プロット、(e) は第三段目の学習におけるオートエンコーダ学習時の誤差推移プロット、(f) は第三段目の交差検定に基づく学習データ(80%)に対する実測値を横軸に予測値を縦軸にとったプロット、(g) は第三段目の交差検定に基づく検定データ(残り20%)に対する実測値を横軸に予測値を縦軸にとったプロット

### 6.5.3 香気化合物の分子パラメータの全般的なオートエンコーダー圧縮に基づくニューラルネットワークを用いた応答予測(93 化合物～OR13a)

図 6-10 に 1500 種類の香気化合物の構造パラメータに対してオートエンコーダーを用いて 125 次元まで圧縮した香気パラメータをもとにして、ニューラルネットワークを用いた応答予測の結果を示す。(a)バックプロパゲーション学習過程におけるエラー推移、(b)トレーニングデータ(予測値-実測値)<sup>2</sup> ヒストグラム、(c) トレーニングデータ予測値-実測値プロット、(d)交差検定(予測値-実測値)<sup>2</sup>ヒストグラム、(e) 交差検定予測値-実測値プロットを示す。オートエンコーダーによるパラメータ圧縮におけるユニット数や学習条件が少しでも変わると学習も、予測もうまくゆかないことがわかった。なお、図 6-10 に示すオートエンコーダーによる次元圧縮のユニット数 125 の時に学習は最も効果的になされているが、その交差検定における予測精度は高くなく、誤差追跡の結果を確認すると、教師データに対する過学習状態となっていると考察できる。

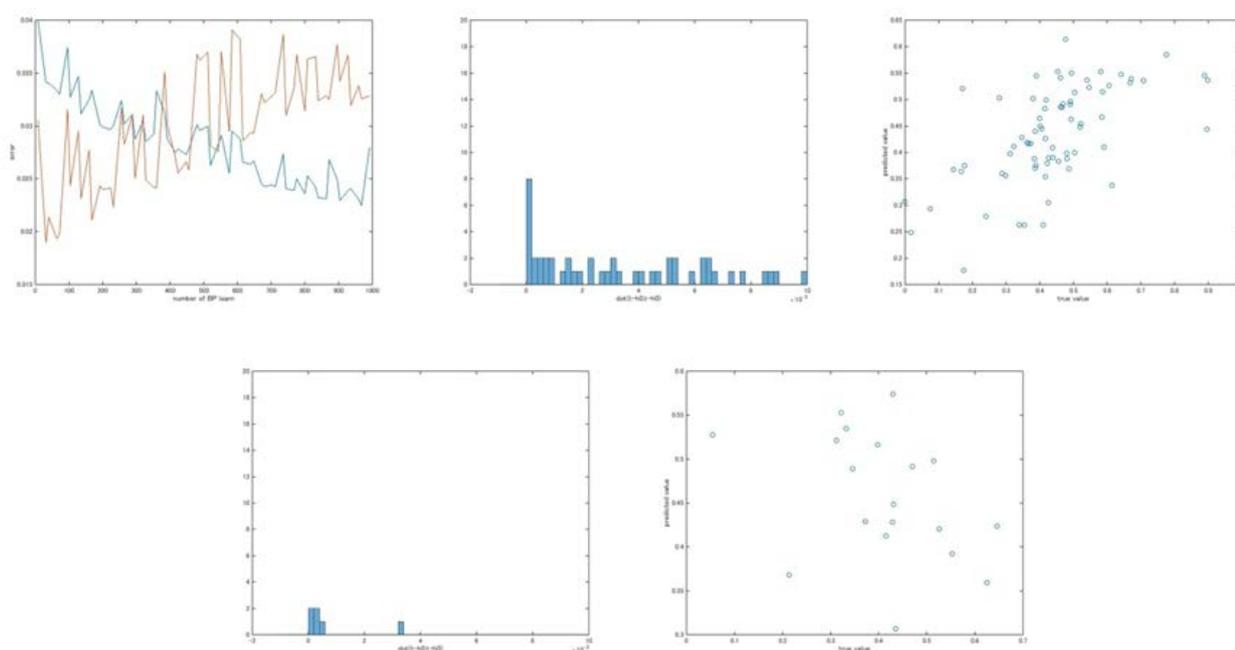


図 6-10. 125 次元まで圧縮した香気パラメータをもとにしたニューラルネットワークを用いた応答予測;(a)バックプロパゲーション学習過程におけるエラー推移、(b)トレーニングデータ(予測値-実測値)<sup>2</sup>ヒストグラム、(c) トレーニングデータ予測値-実測値プロット、(d)交差検定(予測値-実測値)<sup>2</sup>ヒストグラム、(e) 交差検定予測値-実測値プロット

## 6.6 まとめ

この結果をこれまで行っていた条件数を用いた分子パラメータの事前圧縮を行うプロセスと比較したところ、若干の改善が期待できる。

- オートエンコーダーで次元圧縮を行い、圧縮された分子構造パラメータより応答強度の予測ができることを確認した
- バックプロパゲーション学習における条件(ニューラルネットワークの中間ユニット層や学習係数)の調整で予測性能の改善できる可能性がある

- 今後の検討事項としては、さまざまなパラメータを設定する必要があるが、この条件の検討については最低限のものであり、これ以上の詳細な検討はしていない。今後、中間層のユニット数、バックプロパゲーション学習におけるパラメータを再調整する必要があるかもしれない。

## 第七章 結論と今後の課題

古くから匂いの要素が何なのか、というテーマは人々の関心を惹いてきた。現在に至っても自由な遠隔地での匂い再現を目指した技術は少しずつ現実には近づきはしているものの、未完成であり、大きな関心が払われ続けている。本研究は、複数の香りを任意の比率で混ぜ合わせて調合し、すぐに香りを発生させることができる嗅覚ディスプレイを使用する際に、どのような香りの要素を準備したら、様々な香りを提示させられるだろうか、という疑問からスタートした。

匂いは、人体において揮発性化合物の混合物が鼻の粘膜上の嗅覚神経細胞(ORN)にて受容され、それによって生じる感覚刺激の対象である。

匂いを構成する香気物質は 40 万種類もあるとされ、その中には似たような印象のものも多い反面、微量で香気の影響に大きな影響を与える低閾値の香気物質も存在することが知られている。一部の化合物については、人間の感覚器によってガスクロマトグラフよりも鋭敏に検出される反面、分子構造が異なる化合物でも似た匂いとして感じることがあるこのように、生体の嗅覚システムは特定の物質群の差に対しては敏感だが別の物質群の差には鈍感だったりする。したがって揮発性の主に有機化合物の混合物である匂いの差異の評価において、‘構造’を人が感じているように捉えることと、単純な組成を測定する事とは異なると考えられる。匂いを構成する香気物質の混合比率自体は、匂いの印象を表現するには冗長なデータであると考えられ、“感覚量”の空間による解釈が必要である。

本論文では、センサおよび分析器による手法、生理学的・分子生物学的手法、あるいは官能評価によって匂いが受ける多次元の符号化において、得られる情報空間を匂いセンシング空間と呼ぶことにする。

匂いはさまざまな生体内器官において異なる多次元データの形で情報として取り扱われている。もちろん匂いは生物によってのみ情報化されるわけではなく、センサアレイや分析機器によっても情報化することができるし、官能評価も用語と匂いの関係をトレーニングすることが必要だが、ヒトの知覚システム上における多次元情報として取り扱われた結果である。生体内で流通する匂い情報は複合的な ORN 応答情報をもととする複合的な信号である。しかし、ORN による匂いセンシング空間は完全には明らかになってはいない。本検討では、人の嗅覚受容体の応答情報は利用できなかった。ヒトの嗅覚受容体の応答情報の取得は一定の成果を上げているが、ORN 全体の機能に関して検討できるほどの詳細な機能はまだ報告されていない。

他方、知覚上の匂いセンシング空間や脳の高次機能における匂いセンシング空間は経験やトレーニングによってその能力を発達させたりするようである。そのため筆者は複数の香気物質に対する嗅覚神経細胞の応答情報を NMF 法等で解析することが、生体による最初段の“匂いコーディング(符号化)”を明らかにでき、匂いの情報化手法のための基礎的検討につながると考えた。

本研究では、二点の観点から検討を行った。

- ORN による匂いセンシング空間の解析方法
- ORN による匂いセンシング空間の補間方法

ORN による匂いセンシング空間としては、比較的応答情報が得られている昆虫のデータを用いることにした。

筆者は、まず香気データ間の距離尺度に注目した。広いダイナミックレンジを得るのに最適な距離尺度を検討した。次に、生体の感覚量空間に適したセンシング空間として、昆虫の嗅覚受容体の応答データを非負値行列因子分解(NMF : Nonnegative Matrix Factorization)法を用いて解析した。完全な嗅覚受容体による匂いセンシング空間を、最適な距離尺度の選択し、NMF 法などを用いて検討することで、実用的な要素臭の探索に寄与できると筆者は考えている。さらには生体による最初段の“匂いコーディング(符号化)”の解析により、匂いの情報化手法の確立につながると考えられる。生体の嗅覚のセンシング情報を解析して、人工のセンサアレイ系の応答空間が合致させることによりセンサアレイを用いて人工的な嗅覚システムの確立につながると考えられる。

生物学的な受容体の機能解明データの蓄積は近年急速に進んでいる反面、匂いを構成する香気物質に対する、すべての嗅覚受容体の応答を取得することは事実上不可能である。昆虫の ORN 応答データベースでも、限られた化合物に対する応答データしか得られていないことから、受容体の応答を予測することで、ORN による匂いセンシング空間の補間することが必要である。筆者はより多くの香気物質に対して分子構造を考慮した受容体の応答予測の構築するため、香気分子の分子構造パラメータに基づく自己組織化写像(SOM)またはオートエンコーダーと多層パーセプトロンを用いた、嗅覚受容体の応答強度を予測する手法を検討した。2 種類の手法のいずれかを用いることである程度の精度で予測できることが分かった

分子構造を元にした分子構造パラメータ空間を利用することで、多種多様に存在する香気物質に対する匂いセンシング空間との非線形なマッピングを得、匂いセンシング空間を補間する手法は、まだ改善すべき点も多い。だが生体の匂いセンシング空間を補間してゆくことで、その解析に貢献できるものと考えている。

要素臭はまだ明らかになっていないが、以上のように、生体による最初段の“匂いコーディング(符号化)”の解析を得ることで、人の匂い識別特性の理解もすすみ、匂いの情報化手法の確立につながると考えられる。さらには生体の匂いセンシング空間と、人工のセンサアレイ系の匂いセンシング空間が合致させられるようなセンサアレイを設計することで、人工的な嗅覚システムの確立につながると考えられる。

以上

## 謝辞

まず本研究を進めるにあたり、多くのご助言頂きました東京大学先端科学技術研究センター神崎教授並びに神崎研究室の皆様に感謝お礼申し上げます。

今回、専門外の著者が本研究を始めるにあたっては、中本高道教授のおおらかな判断があったからこそ、開始することができました。そして、本研究を進めることができたのは、またこのように修士論文を無事に書き上げることが出来たのは、中本高道准教授の丁寧なご指導の賜物です。本当にありがとうございました。

ポスドクのバートさん、留学生のヨッシリさん、トトさん、研究室においてこまかな質問も丁寧に教えていただいたことで、自分にとって大きな助けとなりました。また早くから機械学習に興味をもって分野に関して調べていた野崎君の MATLAB の m-ファイルがベースになくは 6 章のオートエンコーダーによる検討は、ここに記載したほどには進みませんでした。研究をシェアしてくださったことに感謝します。伊関さんは研究分野の枠を飛び越えるようなテーマについてディスカッションできました。研究室の後輩で、同じように NMF 法に基づく要素臭探索方法の検討を行った本田君と市川さんとは、自分の知っていることをお互い出し合いながら、研究を進めることができました。感謝いたします。ありがとうございました。

秘書の牧山さん、西條さん、今井さん、そして精密工学研究所および電気系の事務の皆さまにも大変お世話になりました。ありがとうございました。

最後に改めて、中本高道教授と中本研究室の皆様に感謝の意を表します。ありがとうございました。

## 参考文献

- [001] M Oz, Y Lozon, A Sultan, KHS Yang, S Galadari, 'Effects of monoterpenes on ion channels of excitable cells', *Pharmacology & Therapeutics*, 2015
- [002] 澁谷 達明 , 市川 眞澄, 「匂いと香りの科学」朝倉書店 (2007/02), ISBN-13: 978-4254102079
- [003] T. Nakamoto and H. P. D. Minh, "Improvement of olfactory display using solenoid valves", *Proc. IEEE Virtual Reality*, 2007, 171-178.
- [004] L. Buck, R. Axel, "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition.", *cell*, 65(1), 175-187, 1991
- [005] 佐藤 幸治, 「昆虫の化学感覚器に発現するイオントロピック型化学感覚受容体」 比較生理生化学 Vol. 29 (2012) No. 2 p. 50-57
- [006] 吉儀 英記, "香料入門", フレグランスジャーナル社 (2002/02)
- [007] Hans Henning "Odor Prism", (1916)
- [008] 斎藤幸子・内山恵二・近江政雄: 「味覚・嗅覚」, p158-185 (2008)
- [009] R. Harper, E. C. Bate Smith and D. G. Land Harper, 'Odour description and odour classification: a multidisciplinary examination', London : Churchill, (1968)
- [009]b J. Robinson (ed) "The Oxford Companion to Wine" Third Edition pg 35-36 Oxford University Press 2006 ISBN 978-0-19-860990-2
- [009]c L. Alley "Wine Sensory Scientist Ann Noble Retires From UC Davis" *Wine Spectator*, August 21st, 2002.
- [009]d San Francisco Chronicle "Ann Noble bio" *Wine Judge*: Accessed Dec. 16th, 2007 Archived 19 December 2010 at WebCite
- [009]e Deuscheweine.de: Das Wein-Aromarad Archived 19 December 2010 at WebCite
- [009]f Owen Bird, *Rheingold - The German Wine Renaissance*, Arima Publishing 2005, pp. 90-97 ISBN 978-1-84549-079-9
- [009]g Adam Jasper and Nadia Wagner, 'Notes on Scent', Issue 32 *Fire Winter* 2008/09, [http://cabinetmagazine.org/issues/32/jasper\\_wagner.php](http://cabinetmagazine.org/issues/32/jasper_wagner.php)
- [010] <http://www.j-ichiyaku.com/kyukaku/t-t.html>
- [011] Moncrieff RW (1967) *The Chemical Senses* (Leonard Hill, London), 3rd Ed.
- [012] amoore(1962)
- [013] ルカ・トゥリン&タニア・サンチェス, 「匂いの帝王」が五つ星で評価する 世界香水ガイド 1437', (2008)
- [014] [http://www.ted.com/talks/luca\\_turin\\_on\\_the\\_science\\_of\\_scent.html](http://www.ted.com/talks/luca_turin_on_the_science_of_scent.html)
- [015] Turin L, 'A spectroscopic mechanism for primary olfactory reception.', *Chem Senses* 21(6):773-791 (1996)
- [016] 肥後 盛秀, '非弾性電子トンネル分光法', *分析化学* 50(10), 637-678, 2001-10-05
- [016]b Adkins, C.J. and Phillips, W.A. (1985) Inelastic electron tunnelling spectroscopy. *J Phys. C: Solid State Phys.*, 18, 1313-1346.

- [016]c Wright, R.H. (1977) Odor and molecular vibration: neural coding of olfactory information. *J. Theor. Biol.*, 64, 473-502.
- [017] E Block, S Jang, H Matsunami, S Sekharan, B Dethier et. al., "Implausibility of the vibrational theory of olfaction", *Proceedings of the National ...*, 2015
- [018] Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R.: *Neuron*, 22, 327-338 (1999)
- [019] Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A., & Axel, R.: *Cell*, 96, 725-736 (1999)
- [020] Kaupp UB, 'Olfactory signalling in vertebrates and insects: differences and commonalities.', *Pubmed*, 2010 Mar;11(3):188-200. doi: 10.1038/nrn2789
- [021] *Nature Neuroscience* 17, 114-120 (2014) / doi:10.1038/nn.3598
- [022] Nakagawa, T., Sakurai, T., Nishioka, T., & Touhara, K.: *Science*, 307, 1638-1642 (2005)
- [023] Nakagawa, T. & Vosshall, L. B.: *Curr. Opin. Neurobiol.*, 19, 284-292 (2009)
- [024] Neuhaus, E. M., Gisselmann, G., Zhang, W., Dooley, R., Strortkuhl, K., & Hatt, H.: *Nat. Neurosci.*, 8, 15-17 (2005)
- [025] Nichols, A. S., Chen, S., & Luetje, C. W.: *Chem. Senses*, 36, 781-790 (2011)
- [026] Pask, G. M., Jones, P. L., Rutzler, M., Rinker, D. C., & Zwiebel, L. J.: *PLoS One*, 6, e28774 (2011)
- [027] Penalva-Arana, D. C., Lynch, M., & Robertson, H. M.: *BMC Evol. Biol.*, 9, 79 (2009)
- [028] Riesgo-Escovar, J., Raha, D., & Carlson, J. R.: *Proc. Natl. Acad. Sci. U. S. A.*, 92, 2864-2868 (1995)
- [029] Sakurai, T., Nakagawa, T., Mitsuno, H., Mori, H., Endo, Y., Tanoue, S., Yasukochi, Y., Touhara, K., & Nishioka, T.: *Proc. Natl. Acad. Sci. U. S. A.*, 101, 16653-16658 (2004)
- [030] Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L. B., & Touhara, K.: *Nature*, 452, 1002-1006 (2008)
- [031] Sato, K. & Suzuki, N.: *J. Exp. Biol.*, 203, 253-262 (2000)
- [032] Sato, K., Tanaka, K., & Touhara, K.: *Proc. Natl. Acad. Sci. U. S. A.*, 108, 11680-11685 (2011)
- [033] "Chapter 3 Identification of Olfactory Receptor Genes from Mammalian (Chiquito J. Crasto (ed.), *Olfactory Receptors: Methods and Protocols*, *Methods in Molecular Biology*, vol. 1003,)" , [http://yosniimura.net/papers/Niimura\\_2013a.pdf](http://yosniimura.net/papers/Niimura_2013a.pdf)
- [034] H. Saito, Q. Chi, H. Zhuang, H. Matsunami, J. D. Mainland, "Odor coding by a Mammalian receptor repertoire", *Science signaling*, 60(2), ra9, (2009)
- [035] Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A. V., Christie, D. L., Chen, C., Newcomb, R. D., & Warr, C. G.: *Insect Biochem. Mol. Biol.*, 38, 770-780 (2008)
- [036] Benton, R., Vannice, K. S., Gomez-Diaz, C., & Vosshall, L. B.: *Cell*, 136, 149-162, (2009)
- [037] Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B.: *PLoS Biol.*, 4, e20 (2006)
- [038] Wicher, D., Schafer, R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H., & Hansson, B. S.: *Nature*, 452, 1007-1011 (2008)
- [039] Abuin, L., Bargeton, B., Ulbrich, M. H., Isacoff, E. Y., Kellenberger, S., & Benton, R.: *Neuron*,

69, 44-60 (2011)

[040] Ai, M., Min, S., Grosjean, Y., Leblanc, C., Bell, R., Benton, R., & Suh, G. S.: *Nature*, 468, 638-640 (2010)

[041] [http://www.nskw.co.jp/analytical/technical/doc/Appnote%20111-Coffee\\_flavor\\_rev2.pdf](http://www.nskw.co.jp/analytical/technical/doc/Appnote%20111-Coffee_flavor_rev2.pdf)

[041]b <https://ja.wikipedia.org/wiki/%E3%82%AC%E3%82%B9%E3%82%AF%E3%83%AD%E3%83%9E%E3%83%88%E3%82%B0%E3%83%A9%E3%83%95%E3%82%A3%E3%83%BC>

[041]c <https://ja.wikipedia.org/wiki/%E3%82%AF%E3%83%AD%E3%83%9E%E3%83%88%E3%82%B0%E3%83%A9%E3%83%95%E3%82%A3%E3%83%BC>

[041]d <https://ja.wikipedia.org/wiki/%E3%83%9E%E3%82%B9%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AB>

[042] [http://people.whitman.edu/~dunnivfm/C\\_MS\\_Ebook/CH2/2\\_3.html](http://people.whitman.edu/~dunnivfm/C_MS_Ebook/CH2/2_3.html)

[043] [https://www.jpo.go.jp/shiryousonota/hyoujun\\_gijutsu/kouryou/mokuji.htm](https://www.jpo.go.jp/shiryousonota/hyoujun_gijutsu/kouryou/mokuji.htm)

[044] [http://www.ogawa.net/study/01\\_01\\_02.html](http://www.ogawa.net/study/01_01_02.html)

[045] 特開 2 0 0 3 - 1 0 7 0 6 7

[046] 特開 2 0 0 7 - 1 6 3 1 9 8 ( P 2 0 0 7 - 1 6 3 1 9 8 A )

[047] <https://ja.wikipedia.org/wiki/%E3%83%9E%E3%82%B9%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AB>

[048] Andrew Dravnieks, 'Atlas of Odor Character Profiles', Astm Intl. (1985), ISBN-13: 978-0803104563

[049] Jane Plailly, Chantal Delon-Martin, Jean-Pierre Royet, "Experience induces functional reorganization in brain regions involved in odor imagery in perfumers", *Human Brain Mapping*, Volume 33, Issue 1, pages 224-234, January 2012

[050] Haruki Takeuchi, Hitoshi Sakano, 'Neural map formation in the mouse olfactory system (Review)', *Cellular and Molecular Life Sciences*, August 2014, Volume 71, Issue 16, pp 3049-3057

[051] <http://molecular-ethology.biochem.s.u-tokyo.ac.jp/molethojp/data/findings/kobayakawa20100414.html>

[052] <http://gara.bio.uci.edu/about/ourSite/index.jsp>

[053] Johnson BA, Woo CC, Hingco EE, Pham KL, Leon M (1999) Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. *Journal of Comparative Neurology* 409: 529-548.

[053]b Leon M, Johnson BA (2003) Olfactory coding in the mammalian olfactory bulb. *Brain Research Reviews* 42: 23-32.

[053]c Johnson BA, Leon M (2007) Chemotopic odorant coding in a mammalian olfactory system. *Journal of Comparative Neurology* 503: 1-34.

[054] M Fisek, "Connectivity and computations in higher-order olfactory neurons in *Drosophila*", 2014

[055] Niimura Y, Nei M (2005) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci USA* 102:6039-6044

[056] Niimura Y, Nei M (2005) Evolutionary changes of the number of olfactory receptor genes in

the human and mouse lineages. *Gene* 346: 23–28

[057] Niimura Y, Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One* 2:e708

[058] Matsui A, Go Y, Niimura Y (2010) Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. *Mol Biol Evol* 27:1192–1200

[059] <http://www.bem.fi/book/>

[060] Hallem et al. (2006) *Cell* 125, 143-160

[061] H. Mitsuno, T. Sakurai, S. Namiki, H. Mitsuhashi, 'Biosensors and Bioelectronics Novel cell-based odorant sensor elements based on insect odorant receptors', *Biosensors and Bioelectronic*, 65, 287-294 (2015)

[062] <https://ja.wikipedia.org/wiki/マイクロプレートリーダー>

[063] [http://www.hamamatsu.com/jp/ja/FDSS\\_uCELL.html](http://www.hamamatsu.com/jp/ja/FDSS_uCELL.html)

[064] T. Weiss, K. Snitz, A. Yablonka, R. M. Khan, Danyel Gafso, Elad Schneidman, and Noam Sobel, 'Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white', *PNAS*, 109 (2012) 49 19959-19964

[065] Jason B. Castro, Arvind Ramanathan, Chakra S. Chennubhotla, 'Categorical Dimensions of Human Odor Descriptor Space Revealed by Non-Negative Matrix Factorization', DOI: 10.1371/journal.pone.0073289

[066] Ohloff, 1986, US4623750

[067] Khan RM, Luk CH, Flinker A, Aggarwal A, Lapid H, et al., 'Predicting odor pleasantness from odorant structure', *J Neurosci* 27: 10015–10023 (2007)

[068] Mandairon N, Poncelet J, Bensafi M, Didier A, 'Humans and Mice Express Similar Olfactory Preferences', *PLoS ONE* 4(1) (2009)

[069] LR Bachtiar, CP Unsworth, RD Newcomb, 'Using Multilayer Perceptron Computation to Discover Ideal Insect Olfactory Receptor Combinations in the Mosquito and Fruit Fly for an Efficient Electronic Nose', 2014

[070] LR Bachtiar, CP Unsworth, RD Newcomb, 'Artificial Neural Network prediction of specific VOCs and blended VOCs for various concentrations from the olfactory receptor firing rates of *Drosophila melanogaster*', in *Medicine and Biology Society (EMBC)*, ..., 2014

[071] Bachtiar LR, Unsworth CP, Newcomb RD, 'Artificial neural network prediction of specific VOCs and blended VOCs for various concentrations from the olfactory receptor firing rates of *Drosophila melanogaster*.', *Conf Proc IEEE Eng Med Biol Soc.* 2014;2014:3232-5. doi: 10.1109/EMBC.2014.6944311.

[072] Bachtiar LR, Unsworth CP, Newcomb RD., "Super E-Noses": Multi-layer perceptron classification of volatile odorants from the firing rates of cross-species olfactory receptor arrays', in *Medicine and Biology Society (EMBC)*, ..., 2014

[073] Amir Madany Mamlouk, Thomas Martinetz, 'On the dimensions of the olfactory perception space', *Neurocomputing*, Volumes 58-60, June 2004, Pages 1019?1025

- [074] Aldrich (Ed.), Flavor and Fragrances Catalog, Sigma Aldrich Chemicals Company, Milwaukee, WI, 1996.
- [075] C.W.J. Chee-Ruiter, The biological sense of smell: olfactory search behavior and a metabolic view for olfactory perception, Ph.D. Thesis, California Institute of Technology, Pasadena, CA, 2000.
- [076] Jordi Fonollosa, Agustin Gutierrez-Galvez, Santiago Marco, 'Quality Coding by Neural Populations in the Early Olfactory Pathway: Analysis Using Information Theory and Lessons for Artificial Olfactory Systems', PLOS one, DOI: 10.1371/journal.pone.0037809
- [077] T.Nakamoto and K.Murakami, "Selection method of odor components for olfactory display using mass spectrum database", IEEE Virtual Reality Conference 2009, 159-162 (2009)
- [078] T.Nakamoto, M. Ohno and Y.Nihe, "Odor Approximation Using Mass Spectrometry", IEEE Sensors Journal 12, 3225 ? 3231 (2012)
- [079] Y.Nihe and T.Nakamoto, "Improvement of Odor Approximation Using Mass Spectrometry", Proc. IEEE Sensors, 2011-2014 (2012)
- [080] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization" Nature 401(1999), 21, 788-791.
- [081] 原田, 中本, 「NMF 法を用いた昆虫の嗅覚受容細胞の応答の解析と近似臭作成方法の研究」, 電気学会全国大会講演論文集, 333-A3 (2013)
- [082] Ethem Alpaydin, 'Introduction to Machine Learning (Adaptive Computation and Machine Learning series) second edition Edition', The MIT Press; second edition edition(December 4, 2009)
- [083] 2012 年度卒業二瓶氏
- [084] 2010 年度卒業村上氏
- [085] 2013 年度卒業本田氏
- [086] 澤田宏, 「非負値行列因子分解 NMF の基礎とデータ/信号解析への応用」, 電子情報通信学会誌 95(9), 829-833 (2012)
- [087] Hallem, E. A., Ho, M. G., & Carlson, J. R.: Cell, 117, 965-979 (2004)
- [088] 印藤 元一, 「合成香料」, 化学工業日報社 (2005)、ISBN-13: 978-4873264608
- [089] Steffen Arctander, 'Perfume and Flavor Chemicals', (1969) ASIN: B000O80GN4
- [090] 2005 年度卒業広田氏の修士論文
- [091] <http://neuro.uni-konstanz.de/DoOR/default.html>
- [092] T. Kohonen. "Self Organization and Associative Memory 3rd edition", Springer-Verlag, Berlin-Heidelberg-New York-Tokio, 1989.
- [093] <http://www.chem-station.com/blog/2012/04/sdf.html>
- [094] <http://www.chemspider.com/>
- [094]b <https://github.com/mcs07/CIRpy>
- [095] 嘉手川氏(東京大学先端科学技術研究センター, 神崎研究室)の卒業論文

## 学会発表など

### (1) 学術雑誌等（紀要・論文集等も含む）に発表した論文、著書

1. 原田 祐希, 本田 克弥, 中本 高道, 「非負値行列因子分解(NMF)法を用いた近似臭作成方法における距離尺度の検討」日本味と匂学会誌, 日本味と匂学会, 20, pp.411-414, 2013 (査読あり)
2. Y. Harada and T. Nakamoto, “Analysis of insect’ s Olfactory Receptor Neuron Response by NMF Method for Odor Approximation”, Sensors and Materials, Vol. 26, No. 3 (2014) 181-187 (査読あり)
3. Yuki Harada, Tomoki Kazawa, Ryohei Kanzaki, and Takamichi Nakamoto, “Response prediction of an insect’ s olfactory receptor neuron by using structural parameters of odorant and Self-Organizing Map”, Sensors Journal, IEEE, 16(3), 580-585, (2016) (査読あり)

### (2) 学術雑誌等又は商業誌における解説、総説（なし）

### (3) 国際会議における発表

1. ○Y. Harada, T. Nakamoto, “Analysis of insect’ s olfactory receptor neuron response by using NMF” ISOEN2013 (International Symposium on Olfaction and Electronic Noses 2013, 査読あり口頭), 2013年7月, Deagu, Republic of Korea
2. ○Yuki Harada, Tomoki Kazawa, Ryohei Kanzaki, and Takamichi Nakamoto, “Response prediction of an insect’ s olfactory receptor neuron by using structural parameters of odorant and Self-Organizing Map (SOM)”, The 15th IUMRS-ICA2014 (International Union of Materials Research Societies - International Conference in Asia 2014, 査読あり口頭), B8-027-003, Fukuoka, Japan
3. ○Yuki Harada, Tomoki Kazawa, Ryohei Kanzaki, and Takamichi Nakamoto, “Response prediction of an insect’ s olfactory receptor neuron by using structural parameters of odorant and Self-Organizing Map”, IEEE SENSORS 2014(査読あり口頭), B2L-F (Paper ID: 1873), 2014年11月, Valencia, Spain
4. ○Yuki Harada, Tomoki Kazawa, Ryohei Kanzaki, and Takamichi Nakamoto, “Response prediction of an insect’ s olfactory receptor neuron by using structural parameters of odorant and Self-Organizing Map”, Digital Olfaction Society 2nd World Congress 2014 (査読あり口頭), ID: 1062, 2014年12月, Tokyo, Japan
5. ○Yuki Harada, ‘Response prediction of an insect’ s olfactory receptor neuron by using structural parameters of odorant for cell-based odor sensor design’, INCF Japan Node International Workshop:Advances in Neuroinformatics 2015(AINI 2015)

### (4) 国内学会・シンポジウム等における発表

1. ○原田 祐希 中本 高道, 「NMF法を用いた昆虫の嗅覚受容細胞の応答の解析と近似臭作成方法の研究」電気学会全国大会（口頭）, セッションNo.333-A3 2013年3月, 名古屋, 日本

2. ○原田 祐希, 本田 克弥, 中本 高道, 「質量分析器データを用いた匂い近似における距離尺度の検討」電気学会 E 部門 総合研究会(ケミカルセンサ研究会, 口頭), CHS-13-007, 2013 年 8 月, 東京, 日本
3. ○原田 祐希, 本田 克弥, 中本 高道, 「非負値行列因子分解(NMF)法を用いた近似臭作成方法における距離尺度の検討」日本味と匂学会第 47 回大会(ポスター, 査読あり), P-135, 2013 年 9 月, 仙台市, 日本
4. ○原田 祐希, 加沢 知毅, 神崎 亮平, 中本 高道, 「自己組織化写像と香気物質の構造パラメータを用いた嗅覚受容神経細胞の応答予測」平成 27 年電気学会全国大会(口頭), 13P-C2, 2015 年 3 月, 東京, 日本

(5) 特許 (なし)

(6) その他(受賞歴等) (なし)

立石科学技術振興財団, 国際交流助成(国際会議発表; 申請課題名「IEEE SENSORS 2014」), 登録番号 2042115



