

エアコン音の聴感印象推定のためのコヒーレンス解析に基づく 脳活動特微量抽出*

☆矢野 肇（神戸大/産総研），滝口哲也，有木康雄（神戸大），
神谷 勝（（株）デンソー），中川誠司（千葉大/産総研）

1 はじめに

新たな音の印象評価手法として，脳活動計測に基づく方法が提案されている^[1-3]．これらの手法は，質問紙法などの心理的手法に比べ，心理的バイアスの影響の低減や意識下の印象評価の可能性といったメリットを有している．

我々はこれまで，エアコン音の聴感印象を対象として，その評価のための脳活動指標の構築に取り組んできた．エアコン音に時間変動を与えた場合の涼しさ，好ましさの印象を一対比較法で推定し，エアコン音聴取時の自発律動の時間的安定性，および事象関連同期／脱同期との関係を調査した^[4,5]．印象と脳活動との間にいくつかの相関を見出したものの，これらの相関は大きくなかった．そこで，脳活動特微量を入力として，エアコン音の聴感印象の尺度値を推定する印象予測モデルを構築し，より頑健なエアコン音の印象評価指標の構築，および印象と関連のある脳活動の抽出にも取り組んだ^[6,7]．Non-negative Tensor Factorization (NTF)^[8]を用いて，エアコン音聴取時の脳磁界の時間一周波数特徴から，その多次元構造を考慮した低次元の特微量抽出を行った後，刺激対の涼しさ，好ましさに関する一対比較判断を教師データとして印象予測モデルを学習した^[6]．しかしながら，一対比較判断の予測に基づいてモデルの評価を行った場合の予測精度は60%に至っていない．現状ではこれらの脳活動指標による印象評価は現実的ではないと言わざるを得ない．

NTFを用いた脳活動特微量抽出の問題点として，近似されたテンソルと元のテンソルとの距離が小さくなるようにフィッティングさせるが，必ずしも印象の予測に有効な特微量が抽出されない点が挙げられる．前稿^[7]では，Discriminant Non-negative Tensor Factorizationを用いて，NTFの距離関数と刺激対に対する優劣の判断の情報を持つ正則化項の和を最小化することで，印象の予測に有効な特微量抽出を試みた．

その一方で，頭皮の外側で観測される脳波や脳磁図には10 Hz程度の律動が多く含まれ，それより高い周波数の成分のパワーは大きくない．時間一周波

数特徴のような周波数成分毎のパワーに基づく特微量にNTFを適用すると，変動の大きな低い周波数成分を表現する基底が得られやすく，脳波や脳磁図の高い周波数成分に聴感印象と関連性があったとしても，その特徴がうまく抽出されない可能性がある．

これまで用いてきた周波数成分のパワーに基づく特微量とは異なる特徴として，大脳皮質機能の領域間の関係性が考えられる．このような関係性はFunctional Connectivity（機能的結合）と呼ばれており，様々な評価指標が提案されている^[9]．本稿では，印象予測モデルの予測精度向上のための新たな特微量として，計測チャンネル間のコヒーレンスに基づく特微量を検討した．全てのチャンネルの組合せに対して計算されるコヒーレンスは，大脳皮質機能の領域間の関係性を反映すると考えられ，その値は各周波数成分のスケールによらない特徴がある．コヒーレンス特微量，およびコヒーレンス特微量と周波数成分のパワーに基づく特微量を組合せた特微量を用いて印象予測モデルを構築し，その予測精度からコヒーレンス特微量の有効性を検証した．

2 脳磁界データ

2.1 刺激音

乗用車内の運転者の着座位置（ヘッドレスト付近）で計測されたエアコン音を基に，線形予測符号化（Linear Predictive Coding: LPC）によってスペクトル包絡を抽出することでモデル化されたエアコン音^[10]を用いて刺激音を作成した．このモデル音に対し，時間変動として正弦波による振幅変調を施し，変調周波数を変化させて7種類の刺激音（0（変調なし），0.2, 0.4, 0.6, 0.8, 1.6, 3.2 Hz）を作成した．なお，刺激音の持続時間は5 s，変調度は0.15に固定し，刺激音圧は被験者毎に，変調周波数0 Hz（変調なし）の刺激音が“はっきりと聞こえ，なおかつちょうど良いと感じる”大きさに設定した．

2.2 計測方法

実験参加の同意を得た聴覚健常者8名（男性6名，女性2名，21-24歳）を被験者とし，刺激音聴取時の

*Brain cortical feature extraction based on coherence analysis for estimating subjective impressions induced by HVAC sound. by YANO, Hajime (Kobe Univ./AIST), TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), KAMIYA, Masaru (DENSO Corp.), NAKAGAWA, Seiji (Chiba Univ./AIST).

脳磁界を計測した。刺激音は予め被験者ごとに7つの刺激音に対する涼しさ、および好ましさの尺度値を一对比較法によって調査し、各印象の尺度値が最大、最小となる刺激音を含む計4つの刺激音を選択した。被験者に4つの刺激のうち2つずつを連続的に呈示し、涼しさ、好ましさの各基準における優劣を評価させた。なお、刺激音は挿入型イヤホンによって被験者の両耳に呈示された。

脳磁界計測は磁気シールドルーム内で122ch全頭型脳磁界計測システム（Neuromag-122™, Neuromag Ltd.）を用いて行われた。全122chで計測された脳磁界信号は通過帯域が0.03–100 Hzのアナログバンドパスフィルタを通過した後、400 Hzでサンプリングされた。

2.3 データの前処理

全122チャンネルで計測されたraw dataから、各刺激呈示を0 msとして、−500–5500 msの区間（エポック）を切り出して用いた。また、不良チャンネルと判断されたチャンネルの信号は用いず、絶対値が3000 fT/cmを超えた信号がいずれかのチャンネルに含まれるエポックも取り除いた。

3 コヒーレンス特徴量

3.1 連続ウェーブレット変換

まず、各エポックのraw dataに対し連続ウェーブレット変換を適用して時間一周波数特徴を抽出した。時間信号 $x(t)$ の連続ウェーブレット変換は次のように表される。

$$W_x(s, \tau) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t - \tau}{s} \right) dt \quad (1)$$

$$\psi(t) = \pi^{-1/4} e^{j\omega_0 t} e^{-t^2/2} \quad (2)$$

ここで、 $*$ は複素共役を表す。 $\psi(t)$ はマザーウェーブレットで、本研究では式(2)の複素モルレーウェーブレットを用いた。パラメータ s, τ はウェーブレットのスケールと時間シフトを表し、これらは解析する周波数と時間に対応する。ウェーブレットのスケールは、分析する周波数区間5–95 Hzを底が2の対数スケールで25分割して得た。

3.2 ウェーブレットコヒーレンス

次に、連続ウェーブレット変換によって得た時間一周波数特徴の0–5000 msの区間を用いてコヒーレンスを計算した。時間信号 $x(t), y(t)$ の間の、スケール s におけるコヒーレンス $C_{xy}(s)$ を次のように定義

する。

$$C_{xy}(s, \tau) = \frac{|T^{-1} \sum_{\tau} W_{xy}(s, \tau)|^2}{P_x(s) P_y(s)} \quad (3)$$

$$P_x(s) = \frac{1}{T} \sum_{\tau} |W_x(s, \tau)|^2 \quad (4)$$

$$W_{xy}(s, \tau) = W_x(s, \tau) W_y^*(s, \tau) \quad (5)$$

ここで、 T は時間平均を取る時の区間長を表す。 $P_x(s)$ はスケログラム $|W_x(s, \tau)|^2$ を時間方向に平均したものであり、スケール s における区間内の平均的なパワーを表している。コヒーレンスは2つ信号のパワーの平均で正規化されているため、信号のスケールによらず、 $0 \leq C_{xy}(s) \leq 1$ の範囲の値を取る。

3.3 低次元なコヒーレンス特徴量の抽出

コヒーレンスはエポック、周波数ごとに、全てのチャンネルの組合せに対して計算することができる。計測チャンネルをノード、コヒーレンスを計測チャンネル間を結ぶ辺の重みと考えると、このグラフは脳皮質におけるネットワークを近似的に表現していると考えることができる。

このグラフの隣接行列をすべての周波数および、エポックにわたって合わせると、4階のテンソル $\underline{C} \in \mathbb{R}_+^{I \times I \times J \times K}$ で表現することができる。ここで、 I はチャンネル数、 J は周波数軸の分割数、 K はエポック数である。近い周波数の隣接行列は類似しており、刺激音の聴取に伴って特定のネットワークのパターンが出現すると仮定することで、周波数、エポック方向の基底を用いて、テンソル $\underline{C}$ を式(6)のようにモデル化した。

$$\underline{C} \approx \hat{\underline{C}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \quad (6)$$

ここで、 $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R] \in \mathbb{R}_+^{I \times R}$ 、 $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_R] \in \mathbb{R}_+^{J \times R}$ 、 $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R] \in \mathbb{R}_+^{K \times R}$ は、それぞれ非負の要素を持つチャンネル、周波数、エポックに関する基底行列、 R はテンソル $\underline{C}$ を表現するランク1成分 $\mathbf{a}_r \circ \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$ の数を表し、“ $\circ$ ”はベクトルの外積（直積）である。式(6)は2つの基底行列を同じものにした時の4階のテンソルのNTFになっており、これは隣接行列が対称行列であるためである。また、NTFの基底行列にはスケールの任意性があるため、行列 $\mathbf{A}, \mathbf{B}$ の各列ベクトルのL2ノルムを1とした。

エポック方向の基底 $\mathbf{c}_r$ の要素 $c_{l,r}$ は、特定の周波数帯における、特定の皮質上の領域間の関係性を示すランク1成分 $\mathbf{a}_r \circ \mathbf{a}_r \circ \mathbf{b}_r$ がどの程度出現しているかを示すと考えられる。したがって、エポックごとの低次元な特徴量として、エポックに関する基底行列 $\mathbf{C}$ の各行ベクトルを用いた。

なお、正確に式 (6) の分解を行えないが、本稿では簡単化のため、2 回出現するチャネル基底 $\mathbf{a}_r$ がそれぞれ異なるものとして Frobenius ノルム規準の NTF のアルゴリズム [8] を用いた。

テストデータから成るテンソル $\mathbf{C}_{test}$ から特徴量 $\mathbf{C}_{test}$ の抽出は、学習データから成るテンソル $\mathbf{C}_{train}$ を分解して得られた基底行列 $\mathbf{A}$, $\mathbf{B}$ を用いて次の式のように行った。

$$\mathbf{C}_{test} = \mathbf{C}_{test}^{(4)}[(\mathbf{B} \odot \mathbf{A} \odot \mathbf{A})^T(\mathbf{B} \odot \mathbf{A} \odot \mathbf{A})]^{-1} \quad (7)$$

ここで、 $\mathbf{C}_{test}^{(4)} \in \mathbb{R}_+^{L \times IJK}$ はテンソル $\mathbf{C}_{test}$ の各要素を、その 4 番目の添字が行の添字になるように並び替えてできる行列で、“ $\odot$ ” は、列のサイズが等しい 2 つの行列に対し列ごとのクロネッカー積をとることを表す (Khatri-Rao 積)。なお、計算の結果負の値となった要素は 0 に置き換えた。

4 評価実験

4.1 評価方法

コヒーレンス特徴量を用いて、被験者毎にエアコン音の涼しさおよび好ましさの印象予測モデルを構築し、その予測精度の評価を行った。また、周波数空間上、および時間一周波数空間上でのパワーを基にした特徴量と、これらの特徴量とコヒーレンス特徴量を組み合わせた特徴量を用いて印象予測モデルを構築し、コヒーレンス特徴量との比較を行った。具体的には次の 5 種類の特徴量間で予測精度の比較を行った: (i) 式 (4) の周波数ごとの平均パワー $P_x(s)$ の平方根を各要素に持つ、周波数、チャネル、エポックの次元を持った 3 階のテンソルに NTF を適用して抽出した特徴量, (ii) 前稿と同様に時間一周波数特徴からなるテンソルに NTF を適用して抽出した特徴量, (iii) コヒーレンス特徴量, (iv) 特徴量 (i) と特徴量 (iii) を連結した特徴量, (v) 特徴量 (ii) と特徴量 (iii) を連結した特徴量。特徴量抽出時に用いる NTF のランク数は $R = 30, 100$ の 2 種類に設定した。なお、全ての特徴量に対して、寄与率が 95% となるように主成分分析で次元圧縮し、白色化を行った。

印象予測モデルはこれまでと同様のモデル [6] を用いた。この印象予測モデルの学習は、刺激対に対応する入力特徴量対の差を、比較判断に対応する 2 つのクラスに分類するように Support Vector Machine の枠組みを用いて行われる [11]。カーネル関数にガウシアンカーネルを選択し、非線形なモデルを構築した。

モデルの予測精度の評価は、まず学習したモデルを用いてテストデータの特徴量から印象の尺度値を予測した後、刺激対に対応する 2 つの予測尺度値の差を計算し、その符号から一対比較判断を予測した。

最後に、脳磁界計測実験において得られた比較判断と予測した比較判断とを比べ、正解率を算出した。また、モデルのパラメータは 5 分割の交差検証で決定した。

4.2 結果・考察

特徴量 (i)–(v) を用いて構築した印象予測モデルによって予測した涼しさ、好ましさに関する一対比較判断の正解率を、8 人の被験者で平均した結果を Table 1, 2 にそれぞれ示す。また、これらの正解率は特徴量抽出時に用いた NTF のランク R ごとに示している。Closed は学習したモデルの評価を学習データで行った場合、Open は学習したモデルの評価をテストデータで行った場合を表している。

涼しさ、好ましさのいずれの評価でも、特徴量 (iii)、すなわちコヒーレンスを含む特徴量が最も一対比較判断の正解率が高い結果となった。また、NTF のランク数が $R = 100$ の場合よりも $R = 30$ の方が平均正解率が高い傾向が見られた。しかしながら、5 つの特徴量の平均正解率の間に有意な差はなく、コヒーレンス特徴量を連結することによる正解率の向上は必ずしも見られなかった。この原因としては、連結した 2 つの特徴量間のスケールの違いや、PCA で得た基底による圧縮では特徴量の変動を十分に捉えることができなかった可能性が考えられる。

次に、コヒーレンスのテンソルに NTF を適用し得られた、チャネルに関する 2 つの基底行列の例を Fig. 1 に示す。本来は同一である基底行列を、異なる 2 つの基底行列として通常の NTF のアルゴリズムで求めているため、同じ基底行列が得られる保証は無い。しかしながら、ほとんどの場合で Fig. 1 のように 2 つの基底行列で同様の基底が得られており、NTF で近似されたコヒーレンスのテンソルの 2 つのチャネル方向での対称性がある程度保たれていると考えられる。

5 まとめ

エアコン音評価のための印象予測モデルの予測精度を改善するために、脳磁界の周波数成分ごとのパワーに基づく特徴量に加えて、計測チャネル間のコヒーレンスに基づく特徴量を検討した。

コヒーレンス特徴量を含む特徴量を用いて学習した印象予測モデルによって、予測した一対比較判断の平均正解率が、涼しさ、好ましさのいずれの場合でも最も高かったが、周波数成分のパワーに基づく特徴量と比較して有意な精度の向上を見出すことはできなかった。印象予測モデルの精度の向上のためには、これらの特徴量をうまく統合するような手法を検討していく必要がある。また、前稿 [7] で提案した手法

Table 1 Mean prediction accuracies of paired-comparative judgment on coolness and their standard deviations (SD) are indicated as follows: Mean $\pm$ SD [%].

Feature	$R = 30$		$R = 100$	
	Open	(Closed)	Open	(Closed)
(i)	58.7 $\pm$ 2.7	(72.4 $\pm$ 18.4)	58.1 $\pm$ 4.3	(84.4 $\pm$ 11.2)
(ii)	59.0 $\pm$ 2.9	(70.0 $\pm$ 13.7)	58.1 $\pm$ 4.3	(91.7 $\pm$ 12.4)
(iii)	57.5 $\pm$ 3.3	(76.8 $\pm$ 16.1)	58.2 $\pm$ 3.9	(83.1 $\pm$ 13.6)
(iv)	59.9 $\pm$ 4.7	(70.5 $\pm$ 14.7)	57.6 $\pm$ 4.1	(80.9 $\pm$ 11.4)
(v)	58.8 $\pm$ 2.8	(71.1 $\pm$ 13.3)	57.6 $\pm$ 4.1	(90.7 $\pm$ 12.0)

Table 2 Mean prediction accuracies of paired-comparative judgment on preference and their standard deviations (SD) are indicated as follows: Mean $\pm$ SD [%].

Feature	$R = 30$		$R = 100$	
	Open	(Closed)	Open	(Closed)
(i)	59.5 $\pm$ 5.1	(69.5 $\pm$ 10.2)	56.4 $\pm$ 2.6	(85.5 $\pm$ 14.9)
(ii)	58.8 $\pm$ 4.5	(70.6 $\pm$ 15.3)	57.7 $\pm$ 3.3	(91.2 $\pm$ 10.4)
(iii)	58.7 $\pm$ 3.2	(77.9 $\pm$ 16.0)	59.9 $\pm$ 3.0	(76.5 $\pm$ 15.5)
(iv)	59.7 $\pm$ 2.3	(64.8 $\pm$ 8.2)	55.4 $\pm$ 2.8	(85.4 $\pm$ 15.0)
(v)	58.8 $\pm$ 4.6	(73.0 $\pm$ 15.5)	58.4 $\pm$ 4.0	(90.2 $\pm$ 9.8)

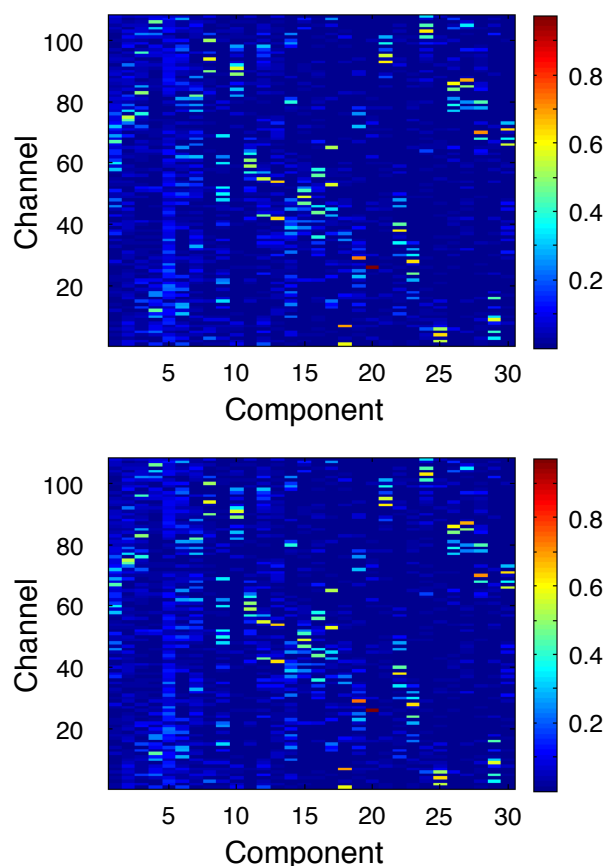


Fig. 1 Examples of two channel basis matrices obtained from a coherence tensor using NTF.

により、刺激対に対する比較判断の情報を取り入れ、コヒーレンスのテンソルから印象の予測に有効な皮質間の関係性を抽出する方法も考えられる。

参考文献

- [1] Y. Soeta *et al.* J.Sound Vib., 258, 419–428, 2002.
- [2] Y. Soeta *et al.*, J.Sound Vib., 277, 479–489, 2004.
- [3] S. Nakagawa *et al.*, ICIC Exp. Lett., 7, 1545–1550, 2013.
- [4] 矢野ら, 音講論 (春), 123–124, 2015.
- [5] 矢野ら, 音講論 (春), 511–514, 2016.
- [6] 矢野ら, 音講論 (春), 539–542, 2017.
- [7] 矢野ら, 音講論 (秋), 573–576, 2017.
- [8] A. Cichocki *et al.*, “Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation,” Wiley, 2009.
- [9] M. Rubinov and O. Sporns, NeuroImage, 52, 1059–1069, 2010.
- [10] T. Hotehama and S. Nakagawa, Proc. of INTER-NOISE 2015, in 15_915.
- [11] R. Herbrich *et al.*, Proc. of AAAI, 80–84, 1998.