

Multilinear Discriminant Analysis を用いた聴感印象推定のための 脳活動特徴量抽出*

☆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が見出されたものの，それらの相関は大きくなかった．

その一方で，脳磁界から抽出した脳活動特徴量を基に，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できた^[6-8]．多チャンネルで計測される脳磁界の時間一周波数表現や計測チャンネル間のコヒーレンスは多次元構造を持ち，テンソルで表現することができる．テンソルの多次元構造を捉えた低次元特徴量を抽出するため，Non-negative Tensor Factorization (NTF)^[9] に基づいた手法をこれまで用いてきた．また，印象の予測に寄与する特徴量を抽出するため，Discriminant Non-negative Tensor Factorization (DNTF) を用いて，NTF による特徴量抽出に刺激対に対する優劣の判断を取り入れた^[7]．しかしながら，これらの特徴量を用いて，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学習した印象予測モデル^[6] の予測精度は 60% 程度にすぎず，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えない．

NTF では元のテンソルを近似したモデルとの誤差の最小化が行なわれる．また DNTF では，抽出される特徴量がクラス毎に分離されるようなペナルティを NTF の誤差関数に付加した損失関数が最小化される．このように NTF に基づく手法では，ある程度モデルとの誤差が小さくなるように基底や特徴量が抽出さ

れるが，このことは印象の予測のための特徴量抽出に必ずしも有効でないと考えられる．例えば，頭皮の外側で観測される脳波や脳磁界には 10 Hz 周辺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周波数が高くなるにつれ減少していく傾向がある．低い周波数帯のパワーを多く含む特徴量の場合，その変動成分を表現する基底が得られやすく，高い周波数成分に聴感印象と関連性があつたとしても，その特徴がうまく抽出されない可能性が考えられる．

本稿では，印象予測モデルの予測精度向上のための特徴量抽出手法として，Multilinear Discriminant Analysis (MDA)^[10] を用いた特徴量抽出を試みた．MDA はクラスラベルが付与されたテンソルからなるデータセットをクラスごとに分離させるような変換を学習し，モデルとの誤差の最小化は行われぬ．比較判断をクラスに対応させて MDA を用いることで，学習データから印象の予測に有用な特徴の抽出が期待できる．脳磁界から計算した周波数特徴量のテンソルに MDA, NTF, DNTF をそれぞれ適用して抽出した特徴量から，印象予測モデルを構築し，その予測精度から MDA による特徴量抽出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^[11]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．

*Brain cortical feature extraction using Multilinear Discriminant Analysis for estimating subjective auditory impressions. by YANO, Hajime (Kobe Univ./AIST), TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), KAMIYA, Masaru (DENSO Corp.), NAKAGAWA, Seiji (Chiba Univ./AIST).

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，刺激音聴取時の脳磁界を計測した．脳磁界計測に先立って，被験者ごとに 7 つの刺激音に対する涼しさ，および好ましさを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，好ましさを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，500–5000 ms の区間をエポックとして切り出した．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．さらに次数 2 の IIR フィルタを用いて電源由来と思われる 60 Hz の成分を除去した．

3 Multilinear Discriminant Analysis を用いた特徴量抽出

3.1 Multilinear Discriminant Analysis

N 個の n 階のテンソルで表されるサンプルからなるデータセットを $\{X_i \in \mathbb{R}^{m_1 \times \dots \times m_n}\}_{1 \leq i \leq N}$ と表す．また，このデータセットは $(n+1)$ 階のテンソルでも表される．このようなテンソルで表されるデータをベクトル化して直接特徴量として用いると，膨大な次元数となり次元の呪いの問題が生じる．また，ベクトル化したテンソルを Principal Component Analysis (PCA) や Linear Discriminant Analysis (LDA) といったベクトルに基づく次元削減手法によって低次元の特徴量抽出を行うことが考えられるが，元のテンソルが持つ多次元的な構造は失われてしまうと考えられる．

Multilinear Discriminant Analysis (MDA) は LDA をテンソルに対して拡張した手法であり，各サンプルがクラスごとに離れるようにテンソルの部分空間を学習する^[10]．MDA でも LDA と同様に，変換後のサンプルの分離度が最大，すなわち，クラス間変動が大きく，クラス内変動が小さくなるような変換を求める．

この最大化問題は次の式で表される．

$$\max_{U_k} \frac{\sum_c N_c \|\bar{X}_c \times_1 U_1 \cdots \times_n U_n - \bar{X} \times_1 U_1 \cdots \times_n U_n\|^2}{\sum_i \|X_i \times_1 U_1 \cdots \times_n U_n - \bar{X}_{c_i} \times_1 U_1 \cdots \times_n U_n\|^2} \quad (1)$$

ここで， $\bar{X}_c$ はクラス c に属するサンプルの平均のテンソル， $\bar{X}$ は全サンプルの平均テンソル， N_c はクラス c に属するサンプル数， c_i は i 番目のサンプルのクラスである．また， $U_k \in \mathbb{R}^{m_k \times m'_k}$, $k = 1, \dots, n$ は行列である． $\times_k$ は k モード積と呼ばれ， $Y = X \times_k U$ は次のように定義される．

$$Y_{i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_n} = \sum_{j=1}^{m_k} X_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n} U_{j, i_k} \quad (2)$$

k モード積はテンソルを k 番目の方向のベクトルの集合と考えた時に，その集合の各ベクトルに行列による線形変換を掛けることを表し，式 (1) では， n 階のテンソルのすべての方向に 1 回ずつ基底行列 U_k による線形変換を施していることになる．

MDA では式 (1) の最大化問題の解となる n 個の変換の基底 $U_1, \dots, U_n$ を求めるが，分離度はすべての基底に関して非線形で，閉じた解を求めることは困難である．そのため，1 つの基底行列に関する最大化を順番に収束するまで繰り返し，すべての基底行列を求める．基底 U_k の 1 回の更新は次のように行う．

$$U_k^* = \arg \max_{U_k} \frac{\text{Tr}(U_k^T S_b^{(k)} U_k)}{\text{Tr}(U_k^T S_w^{(k)} U_k)} \quad (3)$$

ここで，

$$S_b^{(k)} = \sum_{j=1}^{\prod_{o \neq k} m_o} \sum_{c=1}^{m_o} N_c (\bar{X}_j^{c,k} - \bar{X}_j^k) (\bar{X}_j^{c,k} - \bar{X}_j^k)^T \quad (4)$$

$$S_w^{(k)} = \sum_{j=1}^{\prod_{o \neq k} m_o} \sum_{c=1}^{m_o} (X_j^{i,k} - \bar{X}_j^{c_i,k}) (X_j^{i,k} - \bar{X}_j^{c_i,k})^T \quad (5)$$

であり， $\bar{X}_j^{c,k}$ ， $\bar{X}_j^k$ ， $X_j^{i,k}$ はそれぞれ，テンソル $\bar{X}_c$ ， $\bar{X}$ ， X_i の k 番目の方向の添字が行の添字となるように行列に並び替えたときの j 番目の列ベクトルを表す．式 (3) の最大化は LDA のアルゴリズムを用いて解くことができる．

なお，基底 U_k について，元のテンソルの k 番目の方向のサイズ m_k より少ない数の基底が変換に用いられることで，入力されたテンソルよりも小さいサイズのテンソルに変換されるため，MDA によって次元削減を行うことができる．

3.2 比較判断の導入

MDA を用いて，脳磁界から得られた特徴量のテンソルから，刺激対に対する一对比較判断を考慮した

特徴量抽出を行った。まず、比較判断を2つのクラスに対応させるため、比較する2つの刺激うち、先に呈示された刺激を選んだ場合と、後に呈示された刺激を選んだ場合でそれぞれ異なるクラスラベルを付与した。次に、刺激対に対応する2つの特徴量テンソルの変換後の差が、比較判断に対応するクラスごとに分離されるように基底 $U_1, \dots, U_n$ を求めた。刺激対 (A, B) に対応する特徴量テンソルを X^A, X^B とすると、変換後の差は次の式で表される。

$$X^A \times_1 U_1 \cdots \times_n U_n - X^B \times_1 U_1 \cdots \times_n U_n \\ = (X^A - X^B) \times_1 U_1 \cdots \times_n U_n \quad (6)$$

したがって、2つの特徴量テンソルの差 $X^A - X^B$ に2クラスのMDAを適用することによって、比較判断を考慮しながら、元のテンソルよりも小さいサイズのテンソルとして、低次元特徴量を得ることができる。

4 評価実験

4.1 モデルの構築

MDAを用いて抽出した特徴量を用いて、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。

まず、脳磁界のraw dataから切り出されたエポックから、チャンネルごとに短時間フーリエ変換を用いて時間一周波数分析を行った後、振幅成分のみを取り出した。エアコン音は音声などに比べて、比較的定常的で呈示時間が長いため、脳活動特徴の時間変化は少ないと考えられる。そこで、短時間フーリエ変換の各フレームを1サンプルとして扱い、モデルの学習データ数を増加させた。短時間フーリエ変換は長さが500msのハミング窓を用いて、250msごとに計算した。また、すべての周波数ビンのうち4–100Hzに対応する周波数ビンのみを用いた。

短時間フーリエ変換を用いて抽出された周波数特徴量は周波数、チャンネル、試行の方向を持つ3階のテンソルとなる。このテンソルに $n=2$ の場合のMDAを適用し、周波数方向とチャンネル方向の基底 U_1, U_2 を得た。基底による変換によって元のテンソルの周波数とチャンネル方向のサイズが小さくなった3階のテンソルを、試行ごとに切って得られるスライス ($m'_1 \times m'_2$ の行列) をベクトル化し、特徴量とした。なお、周波数基底とチャンネル基底の数は同じにし、基底数は5, 7, 10としたため、特徴量の次元数はそれぞれ25, 49, 100となった。

これまで用いてきたNTFおよびDNTFによる特徴量とMDAによる特徴量を比較するため、周波数特徴量のテンソルに対し、NTF, DNTFを適用してMDAによる特徴量と同じ次元数の特徴量を抽出した。な

お、全ての特徴量に対して、各次元のスケールを揃えるため標準化を行った。

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の枠組みを用いて学習される^[12]。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。

4.2 評価方法

モデルの予測精度の評価は、テストデータから一対比較判断を予測し、脳磁界計測実験において実際に記録された比較判断と比べて、正解率を算出することで行った。具体的には、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。1回の刺激対の呈示に対して、短時間フーリエ変換のフレーム数だけ比較判断が得られるため、全フレームで多数決をとり、その刺激対に対する比較判断とした。なお、モデルのパラメータは5分割の交差検証で決定した。

4.3 結果・考察

NTF, DNTF, MDAによる特徴量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8人の被験者の一対比較判断の平均正解率をFig. 1, 2にそれぞれ示す。

涼しさの評価では、いずれの次元数でもNTFを用いた場合が最も平均正解率が高く、MDAを用いた場合の平均正解率が最も低くなった。好ましさの評価においても、すべての次元数でNTFを用いた場合が最も平均正解率が高かったが、MDAを用いた時の正解率もNTFと同程度の平均正解率となった。その一方で、DNTFを用いた場合の正解率がどの次元数でも最も低くなったが、DNTFのパラメータは涼しさの評価と同様のパラメータを用いており、パラメータ調整によって正解率が上がる可能性があると考えられる。

DNTFとMDAは学習データのサンプルがクラス間で離れるように基底を学習するため、これらの手法は過学習が起こりやすいと考えられる。今回の涼しさの評価において、MDAを用いた場合の正解率の低下は過学習が起こったためだと考えられる。

次に、MDAによって得られた周波数方向の基底 U_1 とチャンネル方向の基底 U_2 の例をFig. 3に示す。この基底が得られたのは涼しさの評価の時であり、印象予測モデルの正解率は70%であった。各基底ベクトルの絶対値が大きい要素がその基底の変換によって取り出される周波数成分またはチャンネルと考えることができる。Fig. 3より、高い周波数の成分が取り出され

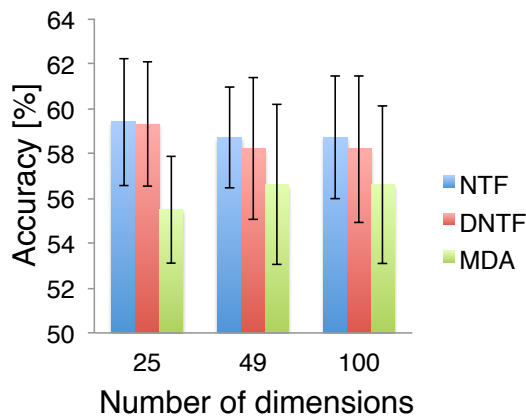


Fig. 1 Mean prediction accuracies of paired-comparative judgment on coolness. Each error bar indicates standard deviation.

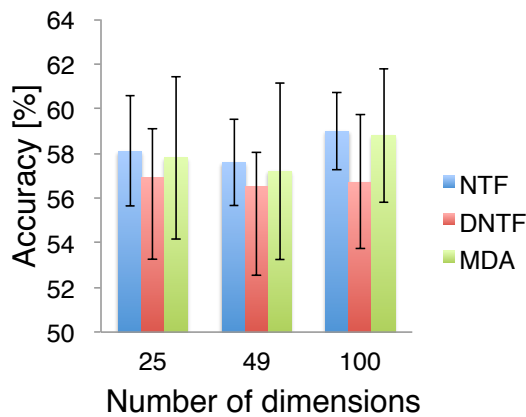


Fig. 2 Mean prediction accuracies of paired-comparative judgment on preference. Each error bar indicates standard deviation.

る基底が多いことがわかる。またチャンネル基底では、近接したチャンネルで高い絶対値を示しており、局所的な領域の特徴を取り出す基底になっていることがわかる。また、これらの周波数成分や領域が涼しさの印象と関連がある可能性が考えられる。

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、MDAを用いて印象の予測に有用な特徴量の抽出を試みた。涼しさの評価ではNTF, DNTFを用いた場合に比べて一対比較判断の正解率が低く、好ましさの評価ではNTFと同程度の正解率であった。一方で、MDAで得られる基底によって、印象と関連のある比較的パワーの小さい脳活動の速い律動成分を取り出せる可能性が示された。印象予測モデルの精度向上のためには、MDAの基底や印象予測モデルの過学習を防ぐ必要があり、例えば正則化項の導入が考えられる。

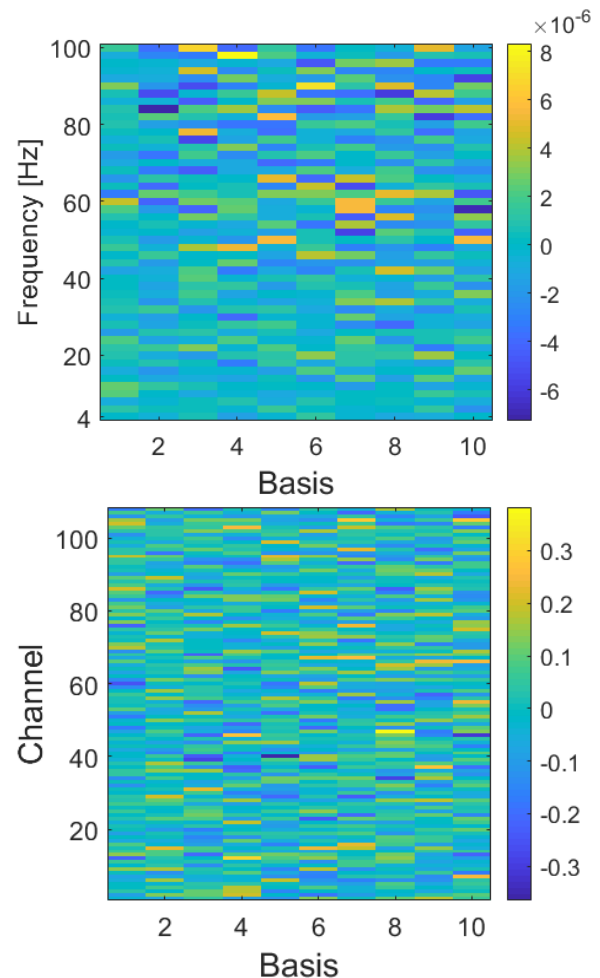


Fig. 3 An example of basis matrices U_1 (top) and U_2 (bottom) obtained by MDA.

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] 矢野ら, 音講論 (春), 755–758, 2017.
- [9] A. Cichocki *et al.*, “Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation,” Wiley, 2009.
- [10] S. Yan *et al.*, IEEE Trans. Image Process., 16(1), 212–220, 2007.
- [11] T. Hotehama and S. Nakagawa, Proc. of INTER-NOISE 2015, in 15.915.
- [12] R. Herbrich *et al.*, Proc. of AAAI, 80–84, 1998.