

脳磁界計測を用いた聴感印象推定の試み —比較判断の重み付けによる精度向上の検討—*

☆矢野 肇 (神戸大/産総研), 滝口哲也 (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, 2, 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对比較法で推定し、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]. その結果、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, 7, 8]. まず、脳活動特徴量として Non-negative Tensor Factorization (NTF) [9] を用いて脳磁界データの多次元構造を考慮した低次元特徴量を抽出した [6]. また、特徴量抽出に刺激対に対する一对比較判断を取り入れ、印象の予測に有効な特徴量の抽出も試みた [7, 8]. 次に、これらの脳活動特徴量と一对比較判断から、印象の尺度値を推定する印象予測モデルを学習した. しかしながら、学習したモデルの予測精度は高々60%程度にすぎず、比較判断を考慮した特徴量を用いてモデルを学習した場合でも、有意な精度の向上は見られなかった. 現状ではこれを脳活動指標とした印象評価は現実的ではないと言わざるを得ない.

特徴量抽出や印象予測モデルの学習に一对比較判断を用いたにも関わらず、印象の予測精度が低い原因の一つとして、一对比較判断自体は2値であるために情報が少なく、たとえ質の低い比較判断が学習データに紛れ込んでいても、モデル学習時にはどの比較判断や試行もすべて同じ重みで扱われてしまう点が挙げられる. 本稿では、印象予測モデルの予測精度向上のために、比較判断にかかる時間(反応時間)を

用いて、比較判断および各データに重み付けを行い、印象予測モデルを学習する方法を提案した. 反応時間が長い場合は、判断が難しかった、もしくは被験者の集中力が落ちていたと考えられ、反応時間は比較判断やその測定データの信頼度を表すと考えられる. この信頼度に応じて特徴量や比較判断を重み付けすることで、より精度の高いモデルの学習が期待される.

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 脳磁界計測実験

実験参加の同意を得た聴覚健常者6名(男性4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. 各被験者について、刺激音の涼しさの評価を行う場合と好ましさの評価を行う場合の計2回の計測が行われた. 計測に先立って、被験者ごとに7種類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、各印象について、その尺度値が最大、最小となる刺激音を含む計4つの刺激音を選択した. 被験者には4つの刺激のうち2つずつを連続的に呈示し、2つ目の刺激の呈示終了後に、各基準における優劣をボタンによって評価するように教示した. 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

*Learning of the predictive model for evaluation of auditory impressions using weighed paired comparative judgment.
by YANO, Hajime (Kobe Univ./AIST), TAKIGUCHI (Kobe Univ.), KAMIYA, Masaru (DENSO Corp.), NAKAGAWA, Seiji (Chiba Univ./AIST).

界信号は通過帯域が 0.03–100 Hz のアナログ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

まず、計 122 チャンネルで計測された raw data に対し、デジタルフィルタを用いて、電源由来の 60 Hz の成分と 2 Hz 以下のトレンド成分を除去した。各刺激呈示を 0 ms として、−500–5500 ms の区間の信号をエポックとして切り出した後、不良チャンネルと判断されたチャンネルの信号と絶対値が 1000 fT/cm を超える振幅が現れるエポックを取り除いた。さらに、以上の処理を施したエポックを時間方向に連結した信号に独立成分分析を適用し、眼球運動に由来する成分や心磁図などノイズと考えられる独立成分を取り除いた。なお、独立成分分析のアルゴリズムには FastICA[11] を用いた。

3 反応時間による重み付けを考慮した印象予測モデルの学習

3.1 反応時間に基づく重み

刺激対の優劣の判断が難しく、呈示の度に判断が入れ替わる場合や、被験者の集中力が低下した状態で得られた比較判断の信頼度は低いと考えられ、このときの特徴量や比較判断を用いて印象予測モデルの学習をした場合、モデルの性能を低下させる恐れがある。その一方で、このような比較判断が難しい場合や、集中力の低下している場合には、刺激対の 2 つ目の刺激の呈示終了から比較判断時のボタン入力までの時間が長くなることが考えられる。したがって、この時間を反応時間と定義すると、これに基づいて各比較判断、もしくは各試行を重み付けし、モデルを学習することで、モデルの予測精度の向上が期待できる。

実際の反応時間の例として、1 回の計測で得られた反応時間のヒストグラムを Fig. 1 に示す。この計測における反応時間の平均は 616 ms、標準偏差は 374 ms であった。Fig. 1 から、多くのボタン入力が 1000 ms 以内にされている一方で、入力までに 2000 ms 以上かかっている場合もあったことがわかる。

反応時間が t のときの比較判断および試行の重みを $w(t)$ とする。Fig. 1 における反応時間の 500 ms 付近のばらつきを考慮すると、反応時間が一定時間以内なら重みを大きく保ち、反応時間が長くなっていくにつれて重みが小さくなるのが望ましいと考えられる。本稿ではこれらの要件を満たす重みとして次の式を用いた。

$$w(t) = \begin{cases} \frac{1}{\sqrt{1+(t/\tau)^2}} & (t \geq 0) \\ 1 & (t < 0) \end{cases} \quad (1)$$

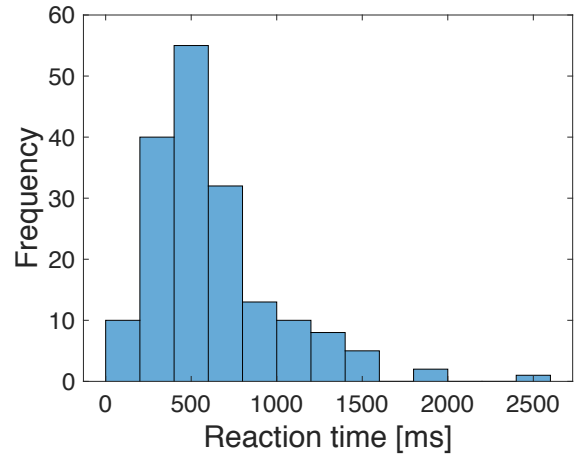


Fig. 1 An example of histogram of the reaction time.

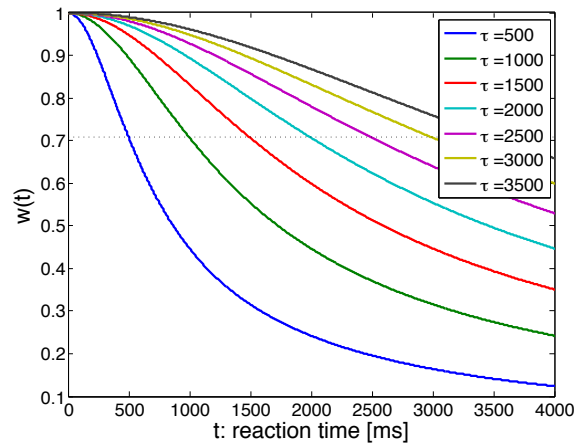


Fig. 2 The weight function $w(t)$ with varying τ .

$t \geq 0$ における式は 1 次ローパスフィルタの振幅応答の式と等価で、ローパスフィルタにおいてカットオフ周波数に対応する τ は重みの減衰の開始を制御するパラメータである。 $t > 0$ において τ を変化させたときの重み $w(t)$ を Fig. 2 に示す。被験者によっては刺激対の呈示終了前にボタン入力がされることがあったため、 $t < 0$ の場合でも重みを定義した。

3.2 重みを考慮した印象予測モデルの学習

本研究では、刺激音聴取時の D 次元の脳活動特徴量 $\mathbf{x} \in \mathbb{R}^D$ から、刺激音の印象の尺度値 $s \in \mathbb{R}$ を予測する印象予測モデルとして次のような線形モデルを考える。

$$s = f(\mathbf{x}) = \langle \boldsymbol{\beta}, \boldsymbol{\phi}(\mathbf{x}) \rangle \quad (2)$$

ここで、 $\langle \cdot, \cdot \rangle$ は内積を表し、 $\boldsymbol{\phi}$ は入力特徴 $\mathbf{x}$ を特徴空間へ写す写像、 $\boldsymbol{\beta}$ はモデルのパラメータである。また、刺激対 (A, B) が与えられたときの比較判断は、対応する特徴量の対を $(\mathbf{x}_A, \mathbf{x}_B)$ とすると、 $f(\mathbf{x}^A)$ と $f(\mathbf{x}^B)$ の大小関係によって決定されると仮定する。

前稿 [8] までは Support Vector Machine (SVM) の枠組み [12] を用いて印象予測モデルを学習したが、この方法では比較判断や試行の重みを考慮してモデルを

学習することはできない．そこで，重みを考慮できるモデルの学習方法として，ロジスティック回帰の枠組みを用いた．特徴量対 $(\mathbf{x}_A, \mathbf{x}_B)$ が観測された時に先に呈示された刺激 A が選択される確率 $p(A | \mathbf{x}_A, \mathbf{x}_B)$ をロジスティックシグモイド関数を用いて次のように表現する．

$$\begin{aligned} p(A | \mathbf{x}_A, \mathbf{x}_B) &= \frac{1}{1 + \exp\{-(f(\mathbf{x}_A) - f(\mathbf{x}_B))\}} \\ &= \frac{1}{1 + \exp\{-\langle \beta, \phi(\mathbf{x}_A) - \phi(\mathbf{x}_B) \rangle\}} \end{aligned} \quad (3)$$

したがって，刺激対 (A, B) に対する比較判断で A を選択したとき 1, B を選択したとき 0 となるような比較判断を表す変数 y と，特徴量対の (写像後の) 差 $\phi(\mathbf{x}_A) - \phi(\mathbf{x}_B)$ が与えられれば，2 クラスのロジスティック回帰と同様にモデルのパラメータ β を学習することができる．

ロジスティック回帰の枠組みを用いた学習に，反応時間に基づく重み付けを導入する方法として，本稿では 2 種類の方法を検討した．1 つ目の方法は，比較判断 y に対して重み付けを行う方法である．反応時間 t に対する重み $w(t)$ を用いて重み付けした比較判断 $\tilde{y}$ は次の式で表される．

$$\tilde{y} = w(t)(y - 0.5) + 0.5 \quad (4)$$

この重み付けによって，小さい重みに対応する $\tilde{y}$ は 0.5 に近づけられる．したがって，反応時間が非常に長い場合はそれぞれの刺激が選ばれる確率が同じとみなされることになる．

2 つ目の方法は，試行に対して重み付けを行う方法である． N 回の試行によって得られた比較判断の集合を $\{y_i\}_{i=1}^N$ ，脳活動の特徴量対の集合を $\{(\mathbf{x}_i^A, \mathbf{x}_i^B)\}_{i=1}^N$ ，反応時間の集合を $\{t_i\}_{i=1}^N$ とする．ロジスティック回帰における対数尤度を重み $\tilde{w}_i$ を用いて次のように重み付けする．

$$\sum_{i=1}^N \tilde{w}_i \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\} \quad (5)$$

ここで， $p_i = p(A | \mathbf{x}_i^A, \mathbf{x}_i^B)$ で， $\tilde{w}_i = \frac{w(t_i)}{\sum_{j=1}^N w(t_j)}$ は和が 1 になるように正規化された反応時間に基づく重みである．これによって，モデルの学習時に反応時間が短い試行を重視し，反応時間が長く信頼度の低い試行が与える影響を小さくすることができる．なお， $\tilde{w}_i = 1/N$ のとき，式 (5) は通常のロジスティック回帰の対数尤度になる．

4 評価実験

4.1 実験条件

被験者ごとに，反応時間に基づく重みを考慮して印象予測モデルを学習し，その予測精度の評価を行った．

まず，前処理後のエポックの刺激呈示後 500–5000 ms の区間において，チャンネルごとに短時間フーリエ変換を行った．短時間フーリエ変換には，長さが 500 ms のハミング窓を用い，時間シフトを 250 ms として計算した．また，4–100 Hz に対応する周波数ビンのみを用いた．刺激音として用い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から得られた脳磁界のパワースペクトルの時系列を時間方向に平均し，その平方根を脳磁界の周波数特徴量とした．

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．このテンソルから NTF を用いて，テンソルの多次元構造を考慮した 50 次元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．

反応時間に基づく重み付けの有効性を検証するため，比較判断と試行のそれぞれに対して重み付けを考慮して学習したモデル（それぞれ wLR1, wLR2）に加え，重み付けなしのロジスティック回帰に基づいて学習したモデル（LR）と，SVM に基づいて学習したモデルで評価を行った．また，いずれのモデルもカーネル法を用いて非線形化し，カーネル関数にはガウシアンカーネルを用いた．

4.2 評価方法

学習されたモデルの予測精度の評価尺度は，テストデータから一対比較判断を予測し，計測において実際に記録された比較判断と比べた時の正解率とした．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．

モデルの性能評価には，10 分割交差検証を用いた．各モデルにはカーネル関数や重み $w(t)$ の減衰開始時間 τ など，いくつかのハイパーパラメータを持つため，分割されてきた学習データセットをさらに 10 分割し，交差検証によって最適なハイパーパラメータを設定した．

4.3 結果・考察

被験者ごとに学習したエアコン音の涼しさ，および好ましさに対する 4 種類の印象予測モデルの平均

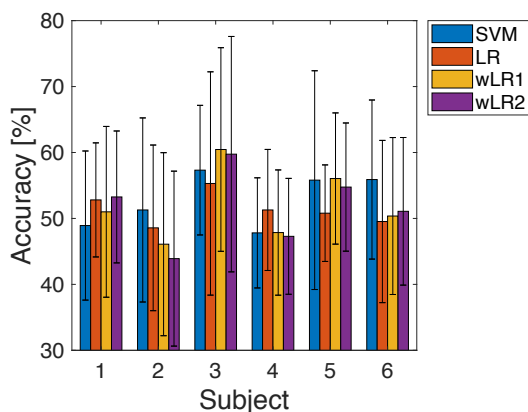


Fig. 3 Mean prediction accuracies of paired comparative judgment on coolness. Each error bar indicates standard deviation.

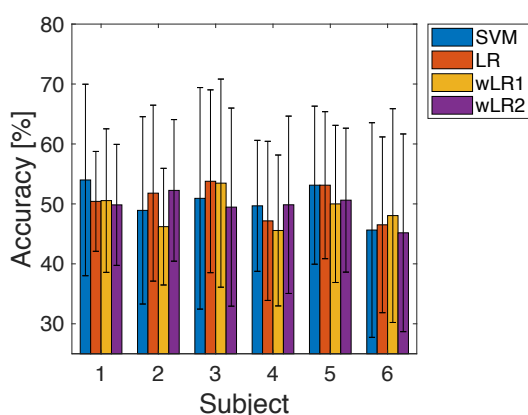


Fig. 4 Mean prediction accuracies of paired comparative judgment on preference. Each error bar indicates standard deviation.

正解率を、それぞれ Fig. 3, 4 に示す。

涼しさ、および好ましさのいずれの評価においても、平均正解率が最も高いモデルは被験者ごとに異なる結果となった。被験者3の涼しさの評価のための印象予測モデルの精度は、反応時間による重みを考慮した場合に60%近くを示しており、他と比べて反応時間が短い時の特徴量から印象の予測に寄与する脳活動パターンを捉えることができたと考えられる。しかしながら、全体的に正解率のばらつきが大きく、各モデルの精度に有意な差は見られなかった。また、今回用いた重みでは、反応時間が0 msから減衰開始時間 τ までに重みが0.7倍になり、 τ 以内の重みが一定というわけではない。今回とは異なる重み関数を用いることでモデルの精度が向上する可能性があると考えられる。

正解率のばらつきが大きいことは、予測精度がデータセットの分割に大きく依存していることを表している。この原因としては、個々の比較判断の質が均一でないことの他に、脳活動特徴量の頑健性の低さが考えられる。特に、刺激呈示終了前に判断のための

ボタン入力がある場合もあったことから、刺激音聴取時の脳磁界に指の運動に関連する反応が重畳し、特徴量に影響を与えた可能性も考えられる。この影響を除く方法の一つとして、計測時に被験者への注意深い教示の他に、刺激呈示終了から一定時間後に判断のセッションを設けることが挙げられる。また、判断セッションにおいてできるだけ早くボタン入力するように教示することで、反応時間に判断の難しさがよりはっきりと反映され、より正確な重み付けができる可能性もある。

5 まとめ

印象予測モデルの予測精度を改善するため、比較判断までの反応時間に基づいて判断の信頼度を示す重みを算出し、重みをモデルの学習に取り入れる手法を提案した。提案法によって、モデルの学習時に、反応時間が長く信頼性が低いと考えられる判断や試行の影響を小さくすることができる。

モデルの性能評価の結果、学習時の重み付けによる有意な予測精度向上は見られなかった。モデルの精度向上のためには、他の重み関数の検討や、刺激音聴取時の脳磁界に影響を与えないような比較判断の計測が必要と考えられる。

謝辞 本研究の一部は、JSPS 科研費 JP18K19820 の助成を受けたものである。

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] 矢野ら, 音講論 (秋), 381–384, 2018.
- [9] A. Cichocki *et al.*, “Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation,” Wiley, 2009.
- [10] T. Hotehama and S. Nakagawa, Proc. of INTER-NOISE 2015, in 15.915.
- [11] A. Hyvärinen and E.Oja, Neural Netw., 13(4-5), 411-430, 2000.
- [12] R. Herbrich *et al.*, Proc. of AAAI, 80–84, 1998.