

## 深層学習を用いた音楽特徴量に基づく感動反応の予測

○森 数馬

((国研)情報通信研究機構)

キーワード：感動，音楽特徴量，深層学習

Deep learning based on music features predict being moved response

Kazuma Mori

(National Institute of Information and Communications Technology)

Key Words : being moved, music features, deep learning

### 目的

音楽は感情の言語と称されるほど、人の感情と密接に関わっている。特に、音楽による感動は、多くの人が求めて止まない感情体験である。喜びや悲しみなどの基本感情と比べ感動研究は数少ないが、楽曲の特徴量から感動を推測することができれば、楽曲推薦サービスや、作曲支援に役立つと期待される。

従来の感動研究では鳥肌感（鳥肌が立つ、背筋がぞくぞくする感覚）のみが検討対象となってきた[1]。しかし、音楽による鳥肌感と涙感の比較を行った研究[2]では、両者が強い快感情を喚起する点で共通するものの、鳥肌感が生理覚醒、涙感が生理鎮静をもたらし、両者は別種の感動であると示唆された。

この研究結果は、感動の鳥肌感と涙感が、別々の音楽特徴量の影響を受けて生起することを示唆する。本研究では、深層学習による検討を行うことで、楽曲の一定区間における音楽特徴量に基づいた感動予測を試みた。

### 方法

心理実験：感動に関する感情ラベルを得るため、音楽聴取実験を行った。実験は54名（男20名、女34名、平均年齢19.1歳[SD = 1.3]）が参加し、個々人の好み3~6曲を聴取した。実験において、参加者は、鳥肌感や涙感が生じたときマウスクリックにより報告した。このクリックは10 Hzで計測を行った。得られた鳥肌の229時点、涙の248時点、ランダムサンプリングした中性238時点感情ラベルとした。予測最適区間の探索：実験でそれぞれの参加者が聴取した合計222曲の音波形に対して音楽情報処理を行い、鳥肌、涙、中性ラベルの前後の時間帯における特徴量を抽出した。音の大きさや高さといった47種類の特徴量についてサンプリングレートを10 Hzに統一して抽出した。どの時間帯の特徴量が最も効果的に3つの感情ラベルを予測するか示すため、感情ラベルの前後10秒間について、平均する時間を1~20秒に変化させて移動平均を取った値で機械学習を行った。深層学習による予測：予測最適区間における音波形に対して周波数帯域を人の聴覚特性に適したメルフィルタバンクで補正した離散フーリエ変換を行うことで、メルスペクトログラムと呼ばれる時間-周波数情報を得た。このメルスペクトログラムに対して、畳み込みニューラルネットワークを行うことで、深層学習による感情ラベル予測を行った。深層学習モデルを構築する際にVGGish[3]を用いた。VGGishはYou tubeから得た持続時間10秒の200万以上の音ファイルを527分類した学習器である。この深層学習器の最終層のみを本研究のデータでfine-tuningし、2層の全結合分類器を接続した上で、softmax関数により予測値を得た。各層の活性化関数はReLUとして、最適化手法にAdamを用いた。また、過学習を抑制するため、ドロップアウトおよびL2 normによる正則化を行った。

### 結果

予測最適区間を探索するため、線形判別分析、サポートベクターマシン、勾配ブースティングなど複数のアルゴリズムを用いて、10分割交差妥当化による機械学習予測を行った。その結果、鳥肌や涙が生じる6秒前から6秒後までの12秒の平均値を用いたカーネル近傍法が最も高い50.1%の正答率を示すとともに、49.4%のF1値を示した(図)。

最も正答率が高かった感情ラベルの付近12秒のメルスペクトログラムを用いて、10分割交差妥当化による深層学習予測を行った結果、65.7%の正答率と65.6%のF1値を示し、同じ区間の従来手法による予測を大きく上回った(図)。

### 考察

本研究から、感動の鳥肌感および涙感を喚起しやすい音楽の時間帯を特定できると示唆された。機械学習による予測を行う際、音楽特徴量を抽出してから従来手法を適用するよりも、メルスペクトログラムに含まれるすべての音情報を使って深層学習による予測を行う方が効果的であると示された。今後の検討において、音楽の一定区間ではなく鳥肌感や涙感を喚起する楽曲そのものを選出する機械学習器の構築が望まれる。

### 引用文献

1. Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A. & Zatorre, R. J. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257-262 (2011).
2. Mori, K. & Iwanaga, M. Two types of peak emotional responses to music: The psychophysiology of chills and tears. *Sci. Rep.* 7, 46063 (2017).
3. Hershey, S. et al. CNN architectures for large-scale audio classification. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 131-135 (2017).

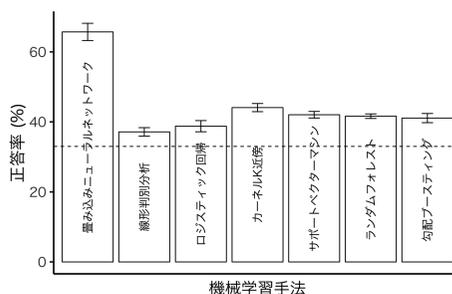


図. 機械学習による感動反応の予測正答率 (破線はチャンスレベル、エラーバーは標準誤差を示す)